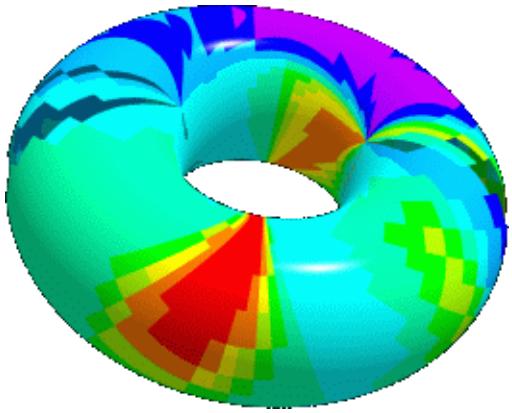


T echniques for
O rganic
R eactions,
V isualization and
S pectroscopy

EROS
Manual

EROS

E laboration of
R eactions for
O rganic
S ynthesis

Version 7

User Manual
(version 0.1)

July 1998

Johann Gasteiger
Susanne Bauerschmidt, Robert Höllering, Wolf-Dietrich Ihlenfeldt, Norbert Karg, Thomas Kostka, Oliver Sacher, Klaus-Peter Schulz, Larissa Steinhauer

Computer - Chemie - Centrum (CCC)
Institut für Organische Chemie
Universität Erlangen - Nürnberg
Nägelsbachstraße 25
D-91052 Erlangen
Germany

Johann.Gasteiger@ccc.chemie.uni-erlangen.de
<name>@ccc.chemie.uni-erlangen.de

http://www2.ccc.uni-erlangen.de

© Copyright Johann Gasteiger

Table of Contents

1 The Philosophy of EROS 5
1.1 Introduction 5
1.1.1 A Knowledge-Based System 5
1.1.2 History of Development 5
1.2 Basic Concepts 7
1.2.1 The Way to Run a Reaction 7
1.2.2 Outline of the EROS System 7
1.2.3 Reaction Generation 8
1.2.4 Evaluation of Reactions 10
1.2.5 Further Development 10
1.3 Reaction Rule File 11
1.3.1 How to Run Reactions: The Rule Header 11
1.3.1.1 Reactors 11
1.3.1.2 Phases 13
1.3.1.3 Modes 15
1.3.1.4 Kinetic Modeling 19
1.3.2 Which Reactions: The Reaction Rules 20
1.3.2.1 Structure Representation 20
1.3.2.2 Reaction Generation 22
2 Operating EROS 25
2.1 General Information 25
2.2 Input Specifications 26
2.2.1 Structure Input 26
2.2.2 Selecting a Rule File 27
2.3 Reaction Generator 28
2.4 Data Output 30
2.5 Visualization of Results 31
3 Sample EROS Runs 35
3.1 The Synthesis of an Amide 36
3.2 Combinatorial Chemistry 39
3.2.1 The Synthesis of Esters and Amides 39
3.2.2 The Synthesis of Pyrazoles 41
3.2.3 1,4-Benzodiazepines 44
3.3 Exploration of all Potential Reaction Products: Chlorination of Benzodioxine 49
3.4 Multistep Laboratory Reaction: Bromination of Phenol 51
3.5 Degradation of Chemicals in the Environment: Degradation of Atrazine and Prometon 54
3.6 Simulation of Mass Spectra 59
3.7 Working with Two Reactors: Acetal Formation from Propanal and Ethanol and Subsequent Simulation of the Mass Spectra 62
4 Writing Your Own Reaction Rules 66
5 Known Limitations 67
6 References 73
7 Appendices 76
7.1 Manual for the CACTVS Editor csed 76
7.2 Manual for the CACTVS Tree Tool cstr 76

Synopsis

EROS (**E**laboration of **R**eactions for **O**rganic **S**ynthesis) is a program system for the simulation of organic reactions. It explores the pathways that given starting materials will follow and attempts to predict the products that will be obtained in those reactions. These predictions are largely built on empirical models that derive their knowledge from the literature, from databases, or from other sources of information.

EROS can be applied to a wide variety of ways to run organic reactions:

- laboratory experiments
- technical processes
- combinatorial chemistry
- degradation of chemicals in the environment
- metabolism of drugs and pharmacokinetics
- simulation of mass spectra

In principle, EROS can be applied to the entire range of organic chemistry, to any reaction type. However, at present not all reaction types have been implemented and for those that have been coded the depth of evaluation, and therefore the quality of predictions, is to different degrees. The ongoing development is mainly directed at improving this situation, extending the types of reactions that EROS can be applied to, and enriching the evaluations of individual reaction types.

In this manual, first, an outline of the main concepts and ideas that have gone into the development of EROS is given in Section 1. The way how to operate EROS is presented in Section 2. Some more details on the use of EROS and its range of applications can be derived from the sample runs that are provided in Section 3. Section 4 describes for the more experienced users or for system operators how to develop reaction rules of their own. No system without limitations and errors, particularly when it is still in active development: The major ones are given in Section 5.

1 The Philosophy of EROS

1.1 Introduction

1.1.1 A Knowledge-Based System

The EROS system (**E**laboration of **R**eactions for **O**rganic **S**ynthesis) simulates organic reactions. It can be used to explore the different reaction pathways that given starting materials can follow to provide products. EROS generates one reaction step after another, thus producing sequences of parallel and consecutive reactions. A major advantage is that in this process EROS exhaustively explores all possible reaction pathways never becoming tired or stopping when it has found one feasible reaction pathway.

It has to be realized that EROS is a knowledge-based system, i.e., it can only generate those reaction types that it knows from. And it can only make decisions between various reaction pathways when rules for the evaluation of reactions have been given to EROS. The evaluation can range from simple rules on which atom and bond types may be involved in a certain reaction type all the way to mathematical functions allowing the calculation of absolute reaction rates.

Clearly, the development of a comprehensive scheme for the evaluation of the entire range of organic chemical reactions is quite an endeavor and will take some time to achieve. The set-up of EROS for the evaluation of organic reactions, however, is hierarchical in nature: starting from rather general rules, a more in-depth evaluation of a particular reaction type can be added later as more knowledge on a specific type of organic chemistry is developed.

The knowledge base defining the kind of chemistry, the kind of organic reactions, EROS can be applied to is kept separate from the program system proper, allowing a flexible development of the chemistry accessible to EROS and easy exchange between different knowledge bases.

1.1.2 History of Development

EROS can look back at a long history of development spanning now more than two decades. The first version was reported in 1978 after five years of development work.[1] Already at that time, we relied on a formal treatment of chemical reactions as electron and bond shifting patterns. The first versions of EROS were applicable to two types of problem cases, to reaction simulation (forward search) and to synthesis design (backward search). The differences in the two types of search problems resided largely in the way the generated chemical reactions were evaluated. With continuing program development and increased sophistication of the system it became clear that the two types of application, reaction prediction and synthesis design, should no longer be handled by a single system.[2] For synthesis design a new approach was taken leading to the development of the WODCA system (**W**orkbench for the **O**rganization of **D**ata for **C**hemical **A**pplications).[3]

From then on, the further development of EROS concentrated on the simulation of the course and products of chemical reactions. A major step was made with EROS version 6 when the knowledge base of EROS was separated from the system proper.[4]

An offspring of EROS was developed for the simulation of mass spectra, MASSIMO (**MA**ss **S**pectra **SIM**ulat**O**r).[5] Work on this specific system eventually convinced us to redesign the EROS system from scratch building on a new representation of chemical structures. Furthermore, new concepts were developed and incorporated into the new EROS version 7 that allow it to handle different ways of running chemical reactions encompassing also the features previously contained in MASSIMO. The EROS 7 version has become operative in the second half of 1998 and has recently presented.[6]

During this long period of development of EROS, the implementation was realized in different programming languages. The first versions had been coded in PL/1, then came FORTRAN 77, whereas version 7 is now coded in the object-oriented-language *C++* with the knowledge base written in the scripting language *Tcl*.

1.2 Basic Concepts

1.2.1 The Way to Run a Reaction

Organic reactions can be carried out under a variety of conditions:
- laboratory synthesis
- technical processes in continuous or batch mode
- combinatorial chemistry
- degradation of chemicals in the environment
- metabolism of nutrients and drugs
- fragmentation and rearrangement of ions in the mass spectrometer

EROS can be applied to all those different ways of running a reaction. This was achieved by introducing new concepts such as reactors, phases, and modes.
These concepts will be explained later in chapter 1.3 in more detail, but some brief remarks seem to be appropriate here. A ***reactor*** can be a three necked flask or a mass spectrometer; it is defined as a place where reactions occur at the same *time*. EROS can handle situations consisting of sequences of reactors in a single program run. For example, one reactor can perform a synthesis, and the second reactor is used to model the mass spectra of the various products obtained in the first reactor.
A reactor can consist of several *locations*, several ***phases***, such as an organic and an aqueous phase with transitions between the phases to be considered. The phases can as well be different compartments of the human body such as the intestinal tract, the blood, the tissue, and the kidney.
There are different ways of how the starting materials are combined in an EROS run which are largely dictated by their *concentration*. At high concentrations the dimerisation of a starting material might occur whereas at very low concentration the starting materials will react only with compounds that have a high concentration such as the solvent or air. These situations are handled by specifying the ***mode*** of a phase.
If enough knowledge on a reaction type is available the evaluation of a reaction can be driven all the way to the calculation of relative or absolute reaction rates. If this is possible, at the end of an EROS run an integration of the partial differential equations automatically derived from the reaction network and the reaction rates will be performed allowing one to make predictions on the development of the various products over the time.

1.2.2 Outline of the EROS System

Figure 1-1 shows the basic outline of the EROS system.

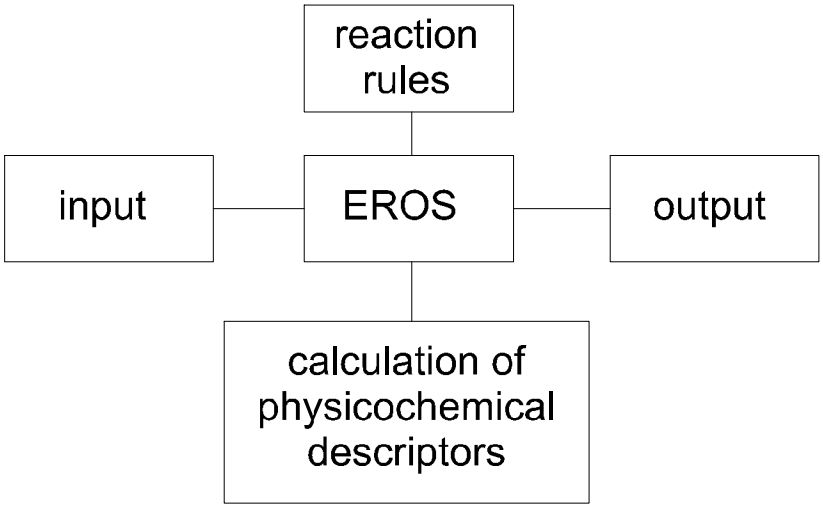


Figure 1-1. Basic outline of the EROS system.

The modules shown in Figure 1-1 are clearly separated from each other communicating through well-defined interfaces. Information between the various building blocks of the EROS system is passed in the form of ASCII files in a format (CTX = clear text) developed in our group quite some time ago.

Information on chemical structures in the CTX format can easily be interconverted with other standard structure exchange formats such as MDL SDfile, SMILES string, JCAMP-CS, SYBYL Molfile etc. Routines for this interconversion are available in the group.

Input to the EROS system can be made by any molecule editor that produces a standard structure exchange format such as the graphical editors *ChemDraw*, the CACTVS editor *csed*, etc. Usually, the EROS system is provided with the CACTVS editor *csed* (see Appendix 7.1).

The results of an EROS run are usually shown as structure diagrams or reaction equations by converting the structure information of a CTX file into a graphical form by programs such as the CACTVS browser *csbr*.

The relationship between individual reaction steps, i.e., which reactions run in parallel and which reaction follow each other can be visualized by the CACTVS tree *cstr* (see Appendix 7.2).

The knowledge base of the EROS system is twofold. One is procedural in nature consisting of a variety of empirical methods for the calculation of physicochemical effects such as heats of reaction, charge distribution, inductive, resonance or polarizability effect.

The other knowledge base gives information on the types of reactions EROS can be applied to. It consists of a header that contains information on the number of reactors, phases, and the modes. This is followed by rules on the various reaction types that will be used in the simulation of reactions by EROS. These rules specify the bond and electron shifting pattern of a reaction type, and the kind of atoms and bonds involved in such an electron reorganization. Furthermore, it can contain methods for the evaluation of such a reaction type: from no evaluation at all (useful for the generation of all possible reactions or isomers) through neural networks for deciding between reaction alternatives, all the way to mathematical functions for the calculation of absolute rate constants.

1.2.3 Reaction Generation

A major characteristic of EROS is that reactions are handled as formal bond and electron shifting patterns. A chemical reaction breaks bonds between atoms and makes new ones. The specific nature of the atoms and bonds involved in this reorganization of bonds make up the large variety of reaction types. This handling of chemical reactions is quite analogous to the

way organic chemists specify reaction mechanisms by drawing curved arrows for the shifting of electrons. An example is the reaction scheme shown in Figure 1-2a

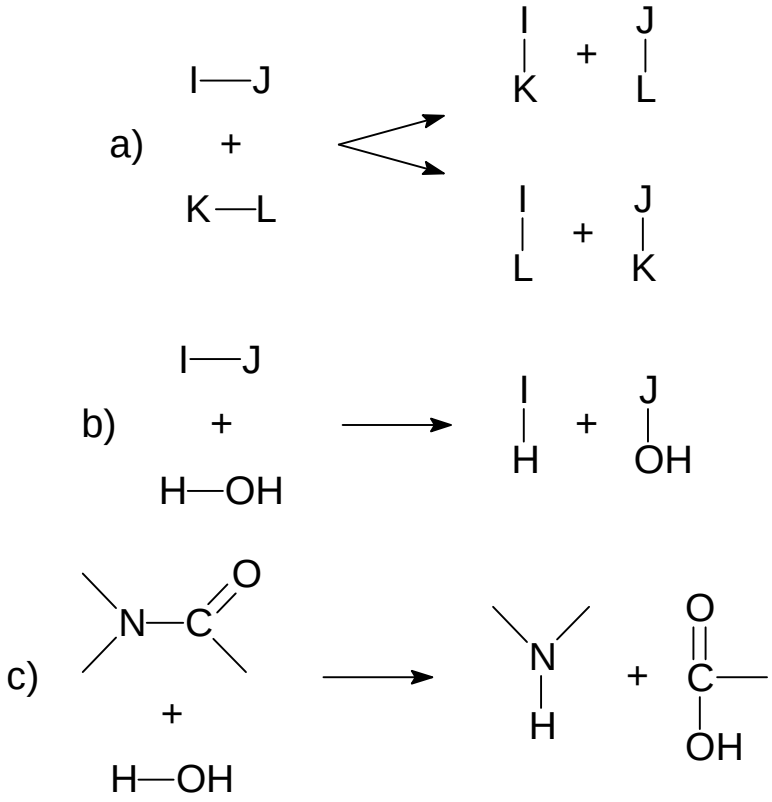


Figure 1-2. A general reaction scheme (a), and two instances: general hydrolysis (b), and amide hydrolysis (c).

In the course of a reaction as shown in Figure 1-2 two bonds are broken and two bonds are made. It is estimated that nearly 50% of all organic reactions follow this scheme: addition to double bonds, elimination reactions, nucleophilic aliphatic substitutions as well as electrophilic aromatic substitutions all break two bonds and make two new bonds. Observe, that nothing is said here about the timing of these events; it can be a concerted reaction or a stepwise process.
When no restrictions are imposed on this reaction scheme both conceivable alternatives for making two bonds will be generated (Figure 1-2a).
In order to define more specific reaction types, constraints on the types of atoms that can be involved in such a bond rearrangement scheme can be imposed. Thus, Figure 1-2b shows the case of a general hydrolysis with atoms K and L now being H and O, respectively, and requiring an additional hydrogen atom to be bonded to the oxygen atom.
Furthermore, rules can be imposed onto which atom will be bonded in the course of the reaction to the hydrogen atom and which to the oxygen atom. Such rules could, for example, be derived from simple electronegativity considerations,
Even more specific reaction types can be obtained when additional restrictions are imposed onto the atoms I and J. Thus, when I is required to be a nitrogen atom and J to be an sp^2 carbon atom having as an additional neighbor a doubly bonded oxygen atom, the case of hydrolysis of an amide is obtained (Figure 1-2c).
All these restrictions can be specified in the rules for a reaction contained in the external reaction rule file (see Section 1.3).

1.2.4 Evaluation of Reactions

The restrictions on the types of atoms and bonds at the reaction center basically are yes/no decisions on whether a reaction can occur or not. More sophisticated evaluation procedures are based on physicochemical properties of the atoms and bonds of the reaction scheme, calculated by rapid empirical procedures.

These methods have been developed over the last 15 years, are described in the literature, and are contained in the program package PETRA (**P**arameter **E**stimation for the **T**reatment of **R**eactivity **A**pplication). Specifically, they involve the calculation of charge distribution,[7][8] estimations of the magnitude of the inductive effect [9], as well as of the resonance stabilization of charges produced on heterolysis,[10] and of the influence of polarizability on charge stabilization. [11] Extensive correlations with physical and chemical data have shown the significance of these calculated values for the physicochemical effects.[12,13,14,15,16,17,18]

These numerical values calculated for the physicochemical effects exerted onto the atoms and bonds involved in the bond rearrangement scheme can then be used for the assignment of a reactivity index, a numerical value for the ease of a reaction to occur. The rule base can contain a mathematical function for the calculation of a reactivity value from the physicochemical descriptors for a certain reaction type. Such functions may have been derived by a statistical analysis of reactivity values for a set of reaction instances. Insertion of the physicochemical values of a given reaction instance into such a mathematical function leads to a specific reactivity value. Instead of explicit mathematical functions, neural networks can be appended to the rule base, both for classifying a reaction as reactive or nonreactive or for calculating a numerical reactivity value.

1.2.5 Further Development

Clearly, the major efforts in the further development of EROS have to go into the extension and refinement of the knowledge base on chemical reactions.

Furthermore, an editor is planned to facilitate the definition of reaction rules.

1.3 Reaction Rule File

The kind of chemistry EROS can handle is laid down in the file that describes the reaction setup and the reaction types that are incorporated into the EROS system. The specific implementations of EROS, and their further development, will largely concentrate on the extension and refinement of this reaction rule file. Therefore, a basic understanding of the concepts and the specific status of the reaction rule file in use is important. System managers and experienced users should have some knowledge on how to interpret the reaction rules in order to be able to develop reaction rules of their own.

In this chapter, we will outline the major features of this knowledge base. Further details are then given in the next two chapters. The reaction rule file is written in the scripting language *Tcl*.[19] It consists of a ***rule header*** that contains information pertinent to the entire EROS run and specifying how reactions are performed, how many reactors or phases are used, etc. This is followed by the ***reaction rules*** specifying the various reaction types that EROS can work with (Figure 1-3). Each reaction type may contain restrictions to limit the scope of its application and procedures for the evaluation of the reactions.

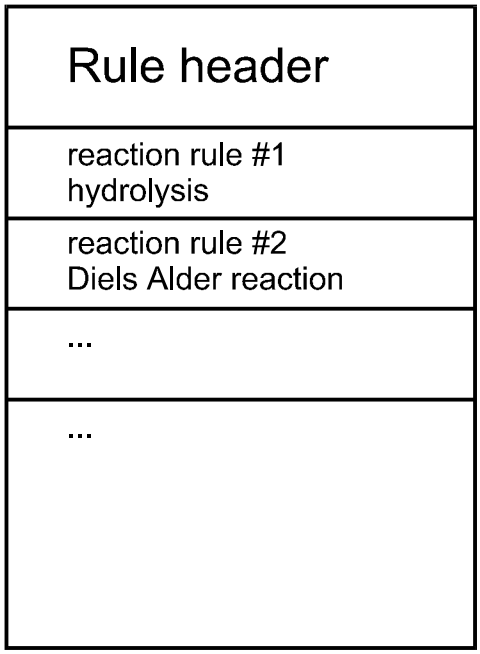


Figure 1-3. Basic set-up of a reaction rule file.

1.3.1 How to Run Reactions: The Rule Header

It has already been said that EROS can be applied to a wide variety of ways for running a reaction, from laboratory synthesis, through combinatorial chemistry to mass spectra simulation.

In order to achieve this, specific concepts have been defined and incorporated into the implementation of the EROS system.

1.3.1.1 Reactors

Definition: A ***reactor*** is a place (vessel, etc.) where reactions occur at the same *time*.

Note, that a reactor is defined by time not just as a physical container. If the way a reaction is run changes, a new reactor has to be introduced. Thus, if a reaction is carried out by adding starting materials over a certain period of time and then the mixture is stirred for an additional period, the system has to be modeled by two reactors, one for the period of addition of the

compounds, and a second reactor for the period of continuous stirring without further addition of compounds.
Two reactors are also needed when an organic reaction is followed by aqueous work-up (Figure 1-4).

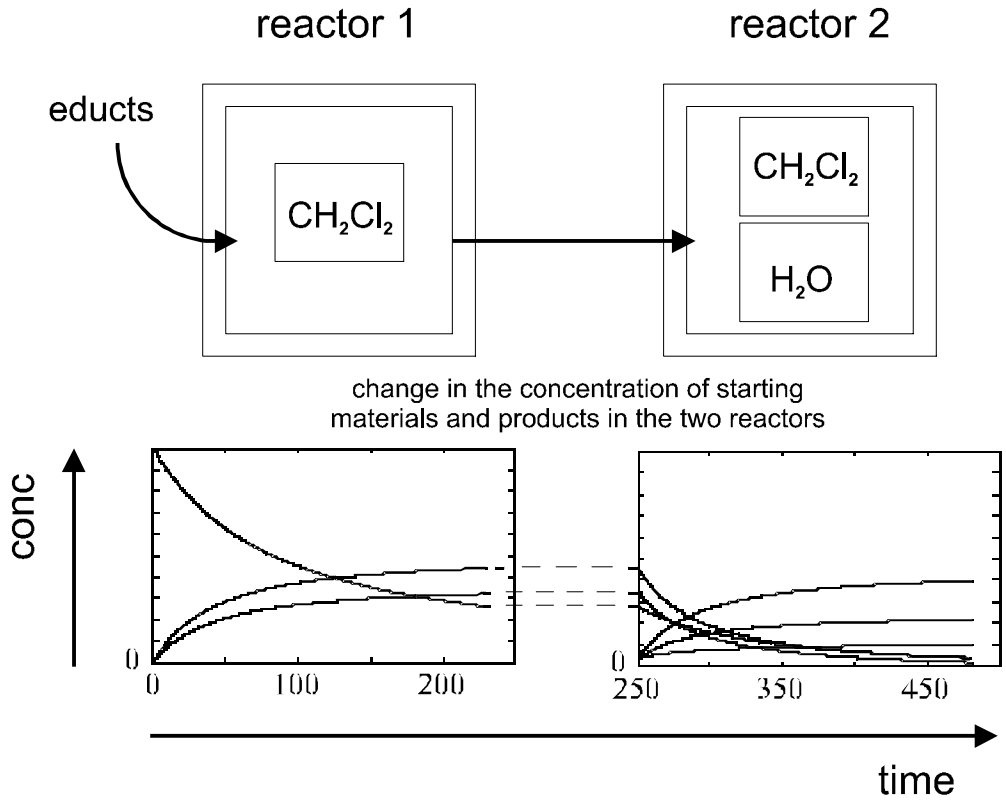


Figure 1-4. The running of a reaction which is followed by an aqueous work-up has to be modeled by two reactors.

Another case for the simulation of reactions by two reactors is given when the products of a reaction are analyzed by a mass spectrometer: The first reactor is used for modeling the reaction, the second reactor for the simulation of mass spectra (Figure 1-5).

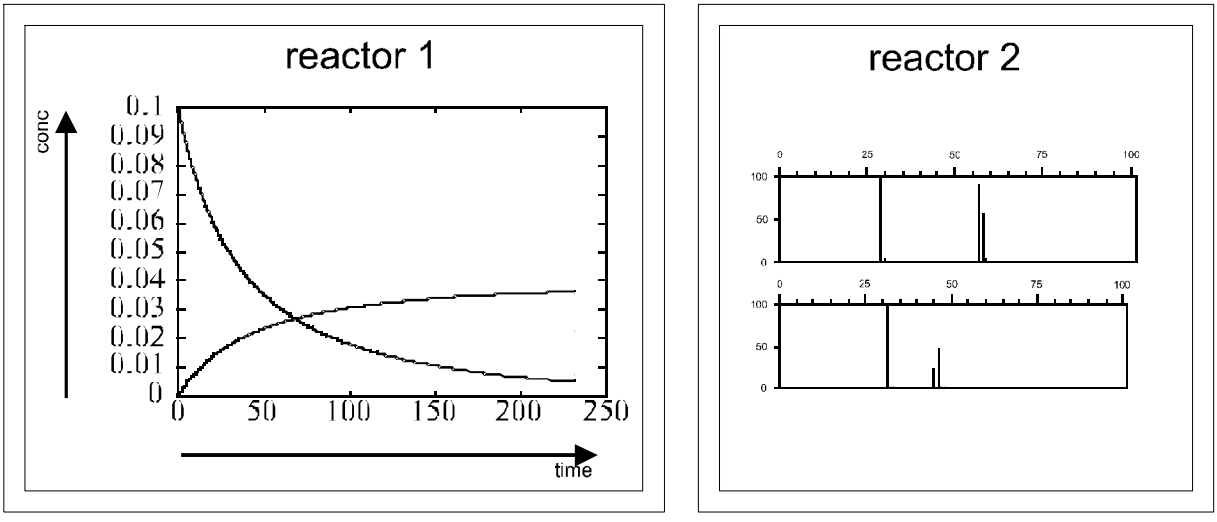


Figure 1-5. A reaction that is followed by GC-MS.

1.3.1.2 Phases

Definition: A ***phase*** is a *place* where a reaction is run that is clearly separated from another such place.

A phase is usually characterized by a homogenous concentration of starting materials. A reactor can consist of one or several phases. In the latter case, transitions between phases have to be considered.

The simulation of a reaction in a flask containing an organic and an aqueous phase has to be modeled by two phases (Figure 1-6). The transfer of each compound between the two phases has to be considered and is handled as a reaction with a rate corresponding to the rate of diffusion.

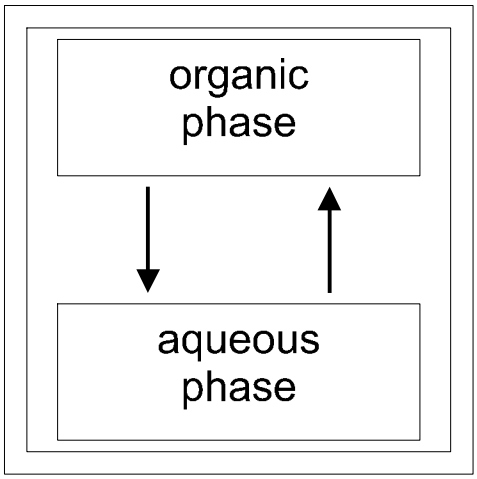


Figure 1-6. A stirred tank reactor consisting of two phases.

Another situation with a reactor consisting of two phases is given when the metabolism of a drug in the blood serum and the subsequent excretion of the drug and its metabolites is modeled (Figure 1-7).

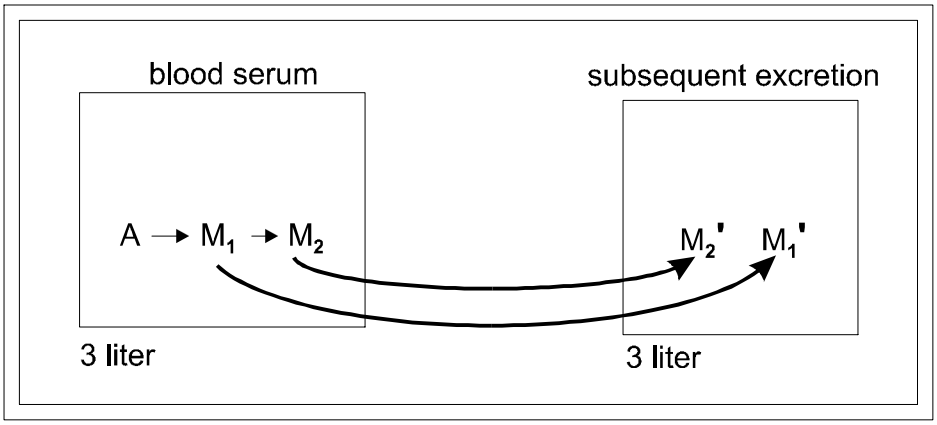


Figure 1-7. The metabolism and excretion of a drug is modeled by a reactor consisting of two phases.

A more elaborate set-up has to be chosen when further details of the events occurring in the distribution and metabolism of a compound in the body should be considered (Figure 1-8). The various compartments of the body important for the pharmacokinetics are represented by phases.

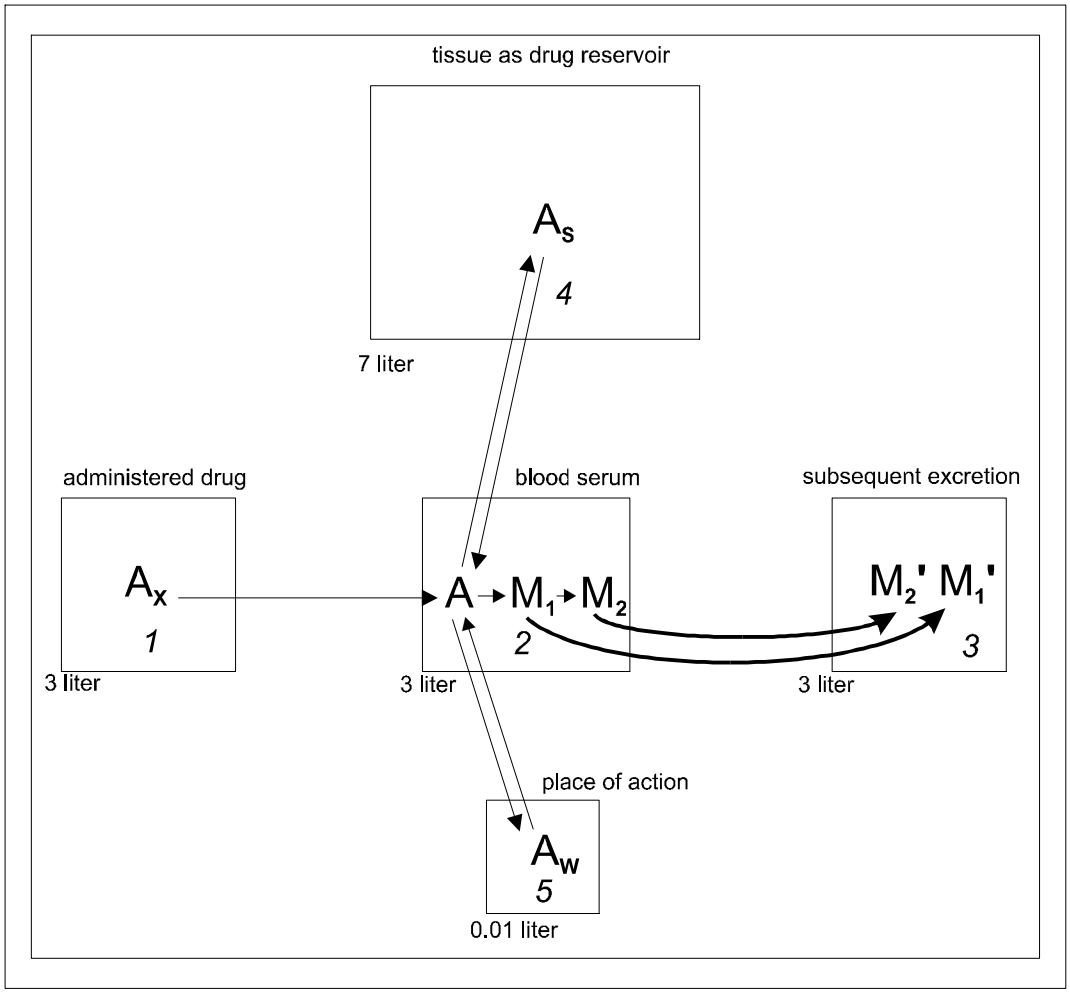


Figure 1-8. The compartments for the pharmacokinetics of a drug as phases and the transitions between them.

A cascade of stirred tank reactors is modeled by one reactor consisting of several phases because the reactions in the various phases are occurring at the same time. Input and output to the phases have to be considered (Figure 1-9).

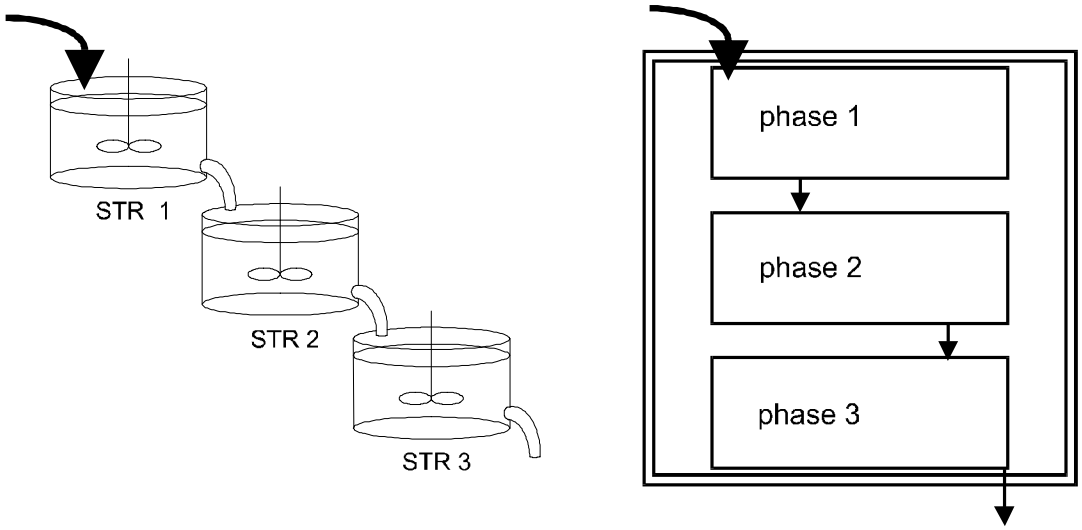


Figure 1-9. A sequence of three (physical) stirred tank reactors (STR) (a) modeled by three phases of a single reactor (b).

Phases play an important role in the modeling of combinatorial chemistry experiments. The various sets of starting materials are assigned to different phases that are specified as having the mode INERT (see section 1.3.1.3) as no reactions are allowed for the compounds assigned to these phases. Basically, these phases are taken as storage devices where single compounds from each set of compounds can be drawn to react with other compounds in subsequent phases. The number of phases in a combinatorial chemistry experiment is given by the number of different sets of starting materials plus the number of reaction steps that have to be performed.
Thus, the combinatorial synthesis of esters from a set of acid chlorides and of alcohols requires three phases (Figure 1-10). The set of acid chlorides is assigned to phase 1, and the alcohols are assigned to phase 2. Then, one after another, one acid chloride is taken from phase 1, an alcohol is taken from phase 2, and both compounds are allowed to react to an ester (and HCl) and are then stored in phase 3.

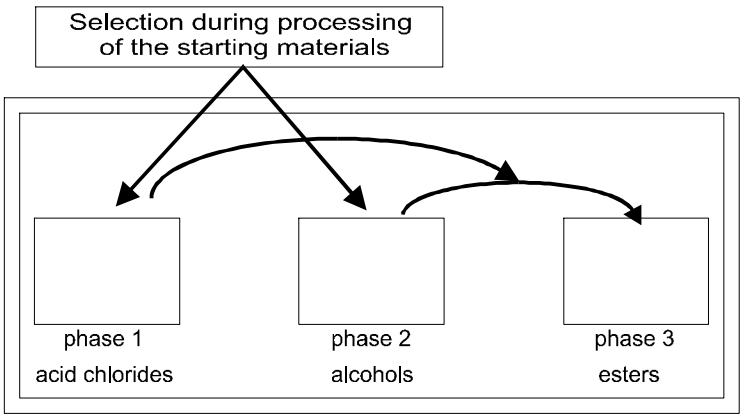


Figure 1-10. The combinatorial synthesis of esters from acid chlorides and alcohols.

The synthesis of tripeptides from activated amino acids (such as esters) and amino acids accordingly has to be handled by four phases (Figure 1-11). The first phase stores the activated amino acids, the second the amino acids. The third phase is used to take dipeptides, and phase 4 to store the tripeptides which result from the reaction of the dipeptides again with activated amino acids.

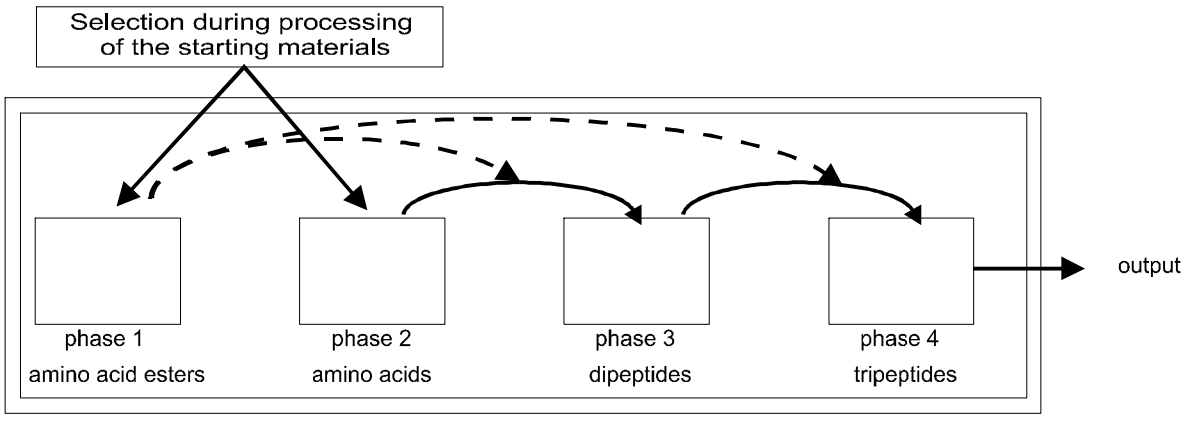


Figure 1-11. The combinatorial synthesis of tripeptides.

1.3.1.3 Modes

The starting materials of a reaction can be combined in a variety of ways that are strongly influenced by the concentration of the species involved. The concentrations govern the kinetic mode, whether monomolecular or bimolecular reactions can occur. Various settings for the

mode of a phase can be specified to take care of this situation and combine the starting materials in the desired fashion. The same is true for all subsequent reaction steps, taking the products of the previous reaction steps as starting materials for the next one and combining them in the fashion specified by the selected mode.
Note, that the combinations of starting materials specified by the setting of the mode parameter will only be explored whether they react with each other. This is not to say that reactions between these combinations of starting materials will indeed be generated. For, the reaction types contained in the reaction rules decide, in the end, whether a reaction is generated. If a combination of starting materials does not contain any of the reaction centers required by the reaction rules, no reaction can be obtained.

Mode = MIX
In this mode, all combinations of starting materials are explored in the generation of reactions. If three starting materials, A, B, and C are given, the following reactions will be investigated:

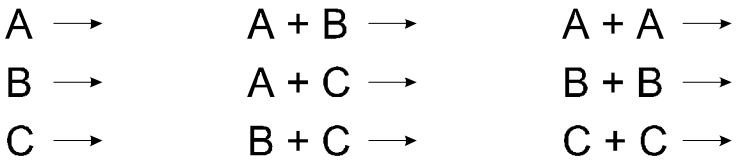


Figure 1-12. Combination of starting materials in the mode MIX.

Clearly, this mode has to be chosen, when the starting materials are given at high concentrations. Note, that no combinations of three starting materials will be investigated as the simultaneous reaction of three molecules is rather unlikely.

Mode = MIX_NO_A_A
With this mode, reactions between molecules of the same sort will not be investigated. Thus, with the three starting materials, A, B, and C, the following combinations (Figure 1-13) will be explored.

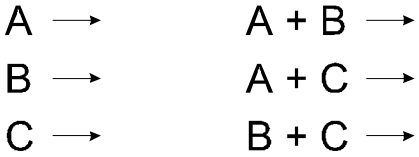


Figure 1-13. Combination of starting materials in the mode MIX_NO_A_A.

This mode comes into play when the concentration of starting materials is at some intermediate value, making the reaction between molecules of the same sort somehow more unlikely.

Mode = MONOMOLEC
In this mode only monomolecular or pseudo-monomolecular reactions will be generated. In the case of three starting materials, A, B, and C, only the following three reactions (Figure 1-14) will be generated.

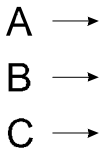


Figure 1-14. Reactions explored in MONOMOLEC mode.

This is the situation with highly diluted solutions. Furthermore, reactions with compounds that are specified as being in high excess such as the solvent, water, oxygen, etc. can be explored. As an example, this mode should be chosen when the degradation of compounds in the environment or the metabolism of a drug is explored. This mode leads to a remarkable speed-up of an EROS run as fewer combinations of compounds and fewer reaction centers have to be analyzed.
Each product of a reaction is individually processed to subsequent reaction steps; no reactions between products are investigated. Thus, a tree of reaction steps will be generated as shown in Figure 1-15.

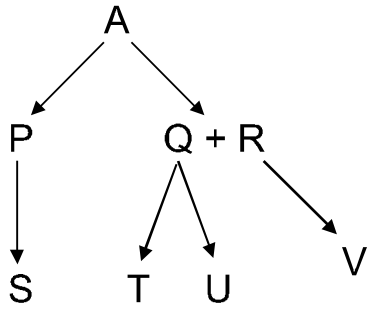


Figure 1-15. Handling of consecutive reactions in the MONOMOLEC mode.

Such a handling of reaction steps is required in the simulation of mass spectra as the high vacuum prevents bimolecular reaction of the products of a fragmentation with each other.
In pseudo-monomolecular reactions, such as the degradation of chemicals in the environment, the reaction steps are handled in an analogous manner, as shown in the following Figure 1-16.

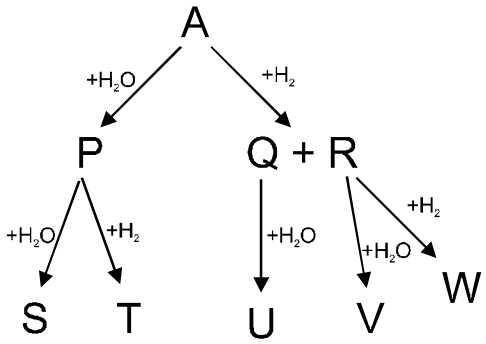


Figure 1-16. Handling of consecutive pseudo-monomolecular reactions.

The scheme shows the fate of a compound A, and its degradation products, P, Q, etc., under hydrolysis and reduction (e.g., reductive dealkylation).

Mode = TUBE
In a laminar tube reactor the products of a reaction are held together and can further react with each other. However, no reactions with the starting materials or the products of other reactions are allowed. The mode = TUBE achieves such a behavior and the following tree of reaction steps is generated (Figure 1-17).

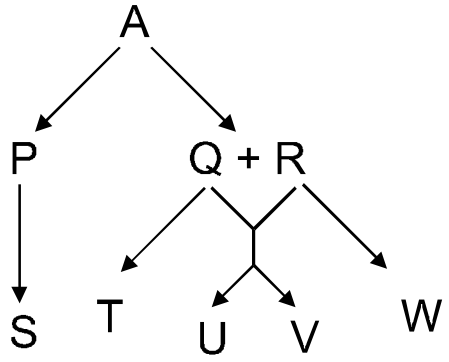


Figure 1-17. Handling of consecutive reactions in the TUBE mode.

Note the difference to the mode MONOMOLEC: in the mode TUBE, the products Q + R may react with each other whereas in the mode MONOMOLEC this possibility is not explored. Another difference is the fact, that with the mode TUBE no kinetics are available.
In the case of a turbulent flow through the tube the reactions and kinetics are the same as for a stirred tank reactor. Then the time for the tank reactor represents the distance in the tube reactor with turbulent flow.

Mode = SURFACE
Reactions can occur at the interface of two phases, one molecule from one phase reacts with one molecule from another phase. This mode is also used for modeling combinatorial chemistry experiments: Two phases each contain a set of molecules; reactions are then generated by drawing one molecule after another from the first set and have it consecutively react with each one of the molecules of the second set in phase 2.

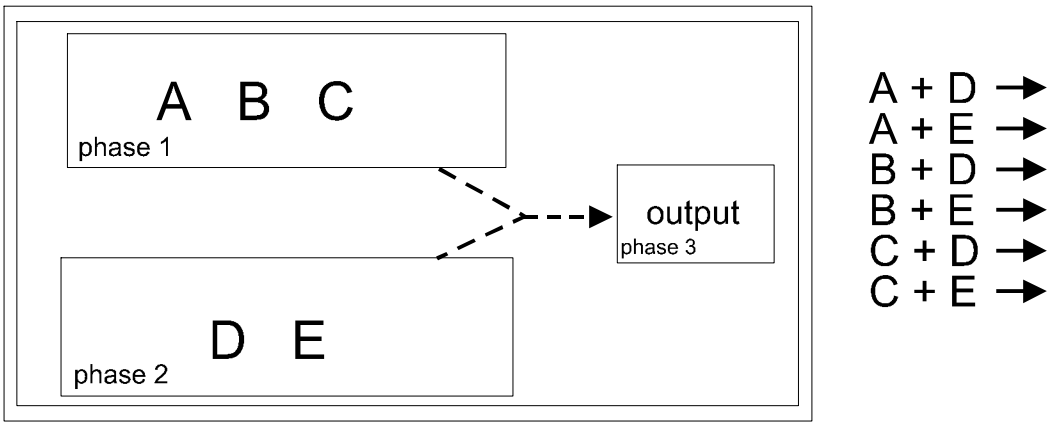


Figure 1-18. Reactions in the SURFACE mode: phase 1 has the mode SURFACE, phase 2 and 3 the mode INERT.

In order to achieve this result, one phase has to be specified with the mode SURFACE, the other as INERT (see below). The results of these reactions are then stored in a third phase.

Mode = INERT
Phases can also be assigned as mode INERT. Then, no reactions are generated in this phase, but such a phase can be used for storing molecules. This feature can be used in combination with a phase having the mode SURFACE for modeling combinatorial chemistry experiments (see also above).
This is explained in Figure 1-19 with the combinatorial synthesis of esters from acid chlorides and alcohols, already mentioned in connection with Figure 1-10.

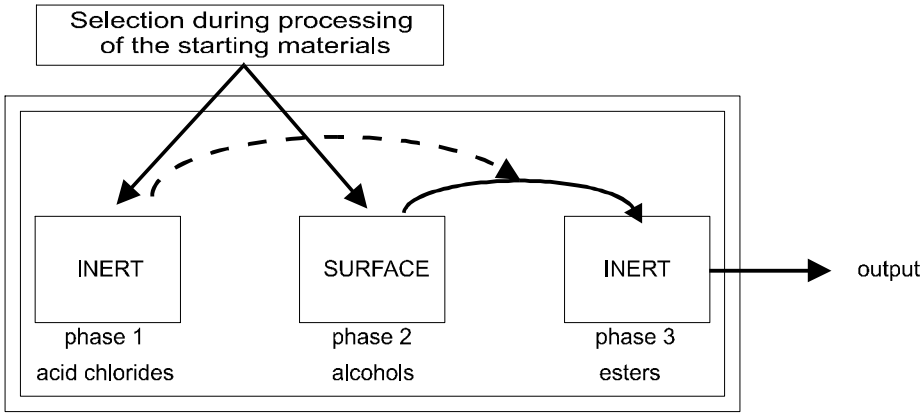


Figure 1-19. The assignment of modes to the phases of a combinatorial chemistry experiment: phase 1 has the mode INERT, phase 2 the mode SURFACE, and phase 3 the mode INERT.

1.3.1.4 Kinetic Modeling

Most organic compounds have a variety of functional groups and, therefore, many reaction pathways are open to ensembles of starting materials. Among these different reactions that reaction will win, will be pursued, that is the fastest one. Therefore, a full modeling of a reacting system should account for the kinetics of the processes.

Clearly, a full kinetic modeling is in most cases beyond our present insight into chemical reactivity. However, by careful analyses of the experimental data and evidence, for quite a few reaction types estimates of relative, or sometimes even absolute [20], reaction rate constants can be achieved.

When evaluation mechanisms for estimating reaction rates are included in the rule files, equations for the rates of different reaction channels are obtained. These partial differential equation can then be integrated to monitor the development of the products over the time (see Figure 1-20).

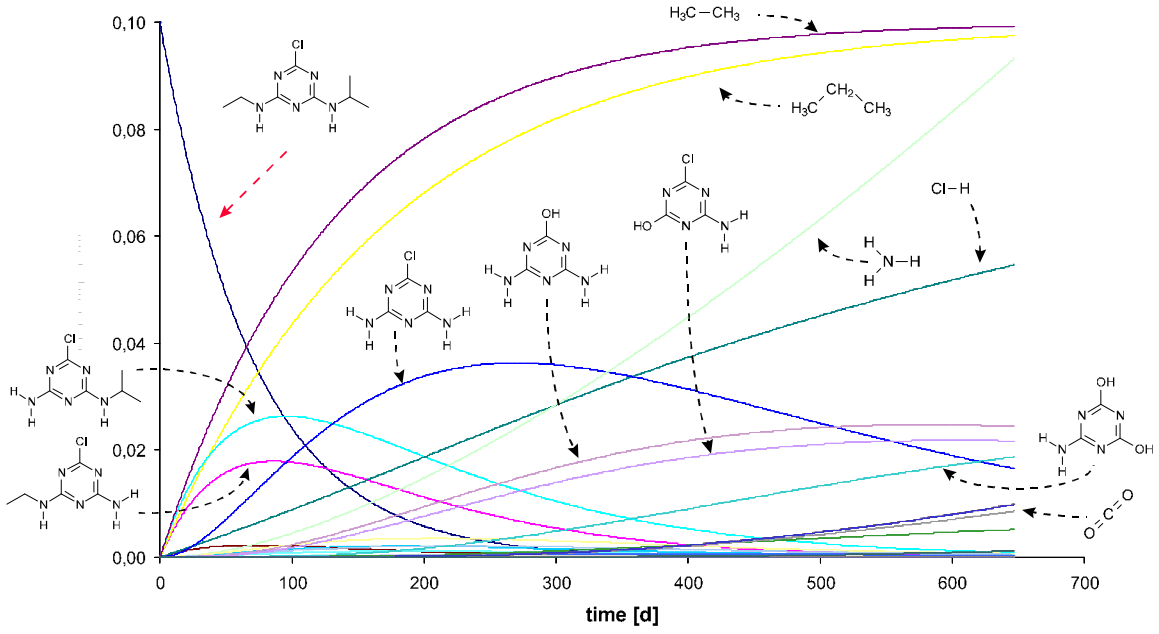


Figure 1-20. The development of products in the degradation of atrazine in soil.

Four different methods are available in the EROS system for the overall evaluation of reaction sequences: the first three for the integration of differential equations, the last one for the evaluation of probabilities of reaction sequences

- the GEAR algorithm [21]
- the Runge-Kutta-method [22]
- the Runge-Kutta-Merson-method
- probability evaluation

The GEAR algorithm is slower than the other two methods but more robust. Usually, it will be the method of choice.

Quite often, the estimation of reaction rates is not possible. As an alternative, probabilities for the different reaction pathways to occur can be given. This is particularly true for the simulation of mass spectra where methods for the evaluation of the probabilities for the different fragmentations or rearrangements of cations and radical cations have been developed.

Based on these probabilities of individual steps, probabilities for entire sequences of steps can be calculated. In the simulation of mass spectra these probabilities are then used for the estimation of peak intensities. The probability kinetics can only be used for monomolecular reactions.

1.3.2 Which Reactions: The Reaction Rules

1.3.2.1 Structure Representation

It has become standard practice to represent chemical structures in the form of a connection table (CT), by lists of the atoms and bonds in a molecule. From the very beginning of the development of EROS we have augmented this information by a list of the free electrons. Thus, we accounted for all valence electrons in a molecule and could also model reactions that involve a shift from free to bonding electrons and vice versa.

Ubiquitous as a connection table is used for structure representation, it should not be overlooked that such a representation also has its limitations. For, in fact, a connection table is basically a valence bond (VB) structure and must fail where a species cannot reasonably well be represented by a single VB structure.

This is true for organometallic structures and for electron deficient molecules such as the boranes. Most organic structures can sufficiently well be represented by a single VB structure; in cases like benzene additional rules can be utilized to take care of cyclic conjugation. However, also in organic species there are situations where a CT fails: A connection table cannot distinguish between a singulett or triplet carbene, cannot handle ionization of a σ-bond, and can only insufficiently represent radical cations.

Just to give an example: The oxygen atom of an enol ether has two free electron pairs, one in conjugation with the double bond, the other orthogonal to it. A connection table cannot distinguish between these two lone pairs. However, it makes quite a difference whether an electron is taken out from the conjugated electron system or from the isolated lone pair.

VB MO

Figure 1-21. A connection table representation of enol ether (VB) augmented with a specification of the number of free electrons cannot distinguish between the two types of orbitals on the oxygen atom (MO).

We became painfully aware of the deficiencies of a connection table in handling radical cations in the course of the development of the MASSIMO system for the simulation of mass spectra. We therefore developed a novel structure representation that overcomes the deficiencies of a connection table.[23]

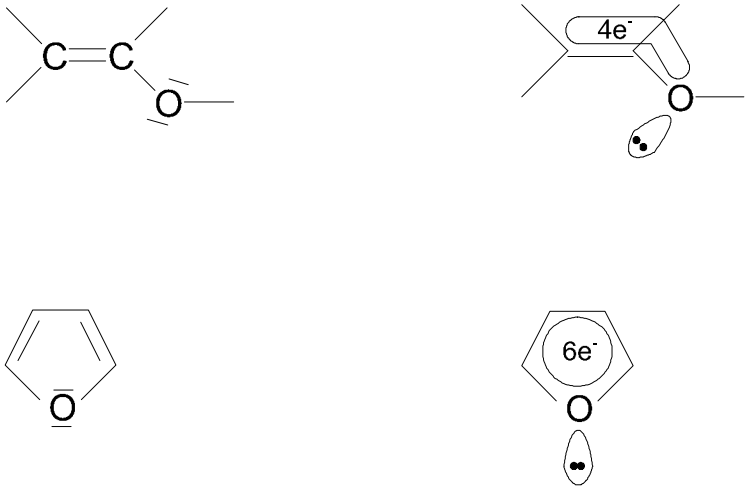
Molecules are handled as species consisting of atoms that are held together by electron systems containing a specified number of electrons distributed over a fixed number of atom centers. Various types of electron systems are handled:

s-systems consisting of two atoms and containing two electrons (normal σ-bonds), σ-systems consisting of two atoms and containing one electron (ionized σ-bond), σ-systems consisting of three atoms and containing two electrons (electron deficient three-center bonds such as those in boranes).

p-systems with one, two, three etc. atoms containing no, one, two, three, etc. electrons, (empty π-orbitals, radicals, free electron pairs, π-bonds, conjugated systems) and *coordinative bonds*.

Details on how this representation can be used for the coding of boranes, organometallic complexes, carbenes, radical cations, etc. can be found in ref.23 and in the Ph.D. thesis of Susanne Bauerschmidt on the internet (http://vermeer.organik.uni-erlangen.de/dissertationen/data/dissertation/Susanne_Bauerschmidt/html).

Two examples, an enolether and furane, are given to explain the new representation form.



connection table (VB structure) MOSES (MO representation)

Figure 1-22. Connection table and MOSES representation of an enol ether and of furane.

With this novel representation the different nature of the two lone pairs on the oxygen atom is distinguished both in an enol ether and in furane. This allows one also to account for two types of ionization at the oxygen atom, from two types of electron systems.
It should be realized that this novel structure representation corresponds more to the description of a molecule by molecular orbitals, hence its name MOSES: Molecular Orbitals: Structures as Electron Systems.
The entire EROS system version 7 is founded on the novel MOSES representation which has been implemented in the object oriented language *C++*. However, interconversion routines have been incorporated into the EROS system to also access it by a traditional connection table. For, a VB structure is quite a reasonable representation for most organic structures.
Thus, although all internal structure manipulations, in particular the generation of reactions, are made on the MOSES data structure, reaction schemes can be specified as bond and free electron shifting patterns in a VB notation or as changes in electron systems in a MOSES notation.

1.3.2.2 Reaction Generation

As has been detailed in the previous section, reactions can be specified as patterns of bond and free electron shifting schemes changing the connection tables of the starting materials (Figure 1-23) or as changes in electron systems working on the MOSES representation as indicated in Figure 1-24.

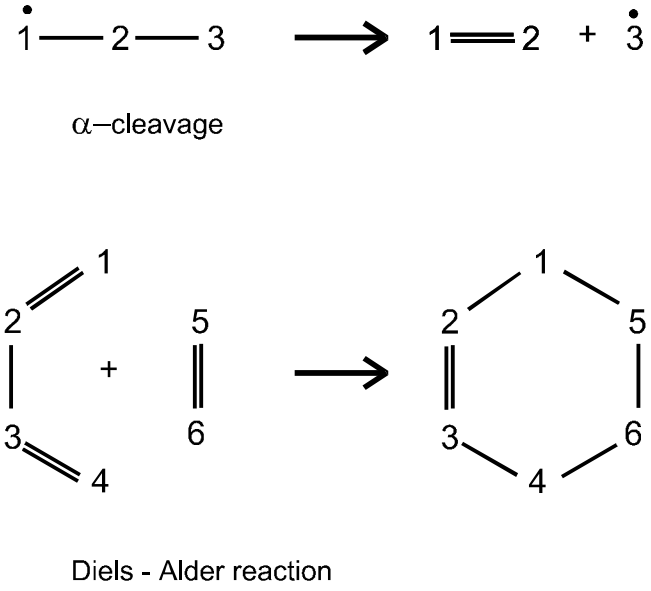


Figure 1-23. Reactions described as changes in the connection tables of the ensemble of starting materials.

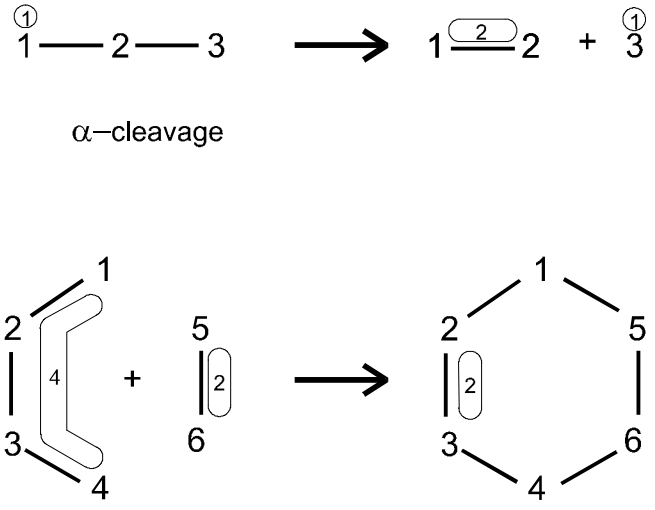
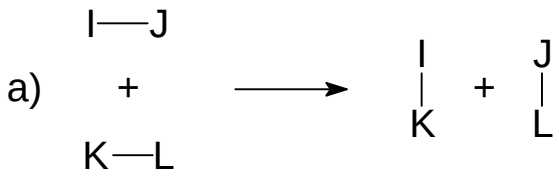


Figure 1-24. Reactions described as changes in the electron systems of the starting materials.

A major characteristic of the EROS system is that reactions are handled in a formal manner, as electron shifting patterns. Organic reactions cover a limited number of such shifting schemes; it is the different nature of the atoms and bonds involved in those reaction schemes that account for the large variety of reaction types.

The atoms and bonds involved in the electron rearrangement make up the reaction center, or, as it is called in the EROS system, the reaction substructure. The reaction rule has to specify the atoms and bonds (or electron systems) that are part of the reaction substructure. First, the number of atoms of the reaction substructure have to be given and how they are bonded to each other. Then, constraints on the nature of the atoms may be given, either as a list of specific atoms, such as O, N, S, Cl, Br, or I, or as individual atoms only, such as only C. Furthermore, an atom may also be restricted to be at a certain hybridization state, such as C sp^2. Constraints can also be given for the bonds. Thus, it may be required that a bond will be broken only if it is part of a multiple bond.

Constraints can be specified not only for the atoms and bonds of the reaction substructure (reaction center) but also for the atoms bonded to the reaction substructure and the bonds between those atoms. Figure 1-25 shows reaction types that have more and more restrictions and thus become more and more specific.



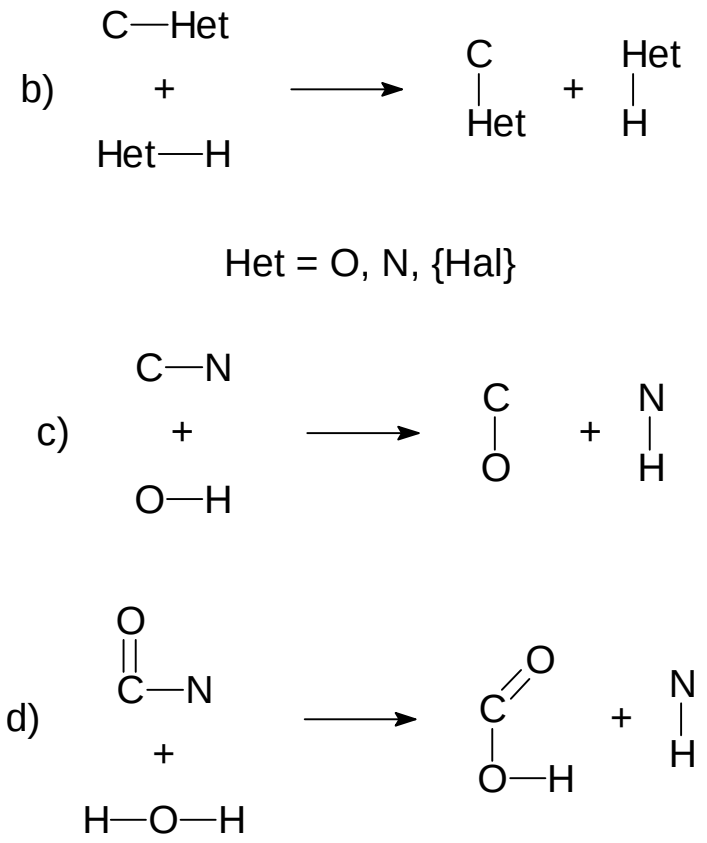


Figure 1-25. A very general reaction scheme (a) becomes more and more focused as more specific restrictions are imposed onto the reaction substructure and its neighborhood.

The reaction scheme of Figure 1-25a is very general breaking any combination of two bonds. Clearly such a reaction scheme should only be used in exceptional cases (such as the generation of isomers starting from a given molecule) as usually for too many reactions including many unreasonable ones will be generated.

The restrictions on the types of atoms shown in Figure 1-25b for the atoms I, J, K, and L of Figure 1-25a lead to a reaction type that, among others, covers all nucleophilic substitutions (aliphatic and aromatic). The reaction type of Figure 1-25c is even more specific until constraints onto atoms adjacent to the reaction substructure (Figure 1-25d) limit the reaction type to the hydrolysis of amides.

2 Operating EROS

2.1 General Information

This graphical user interface was designed to facilitate communication between the user and the EROS7 program for reaction prediction. Instead of repeatedly typing long file names and more or less cryptic command line options or having to develop shell scripts, an intuitive access to EROS7 is provided.

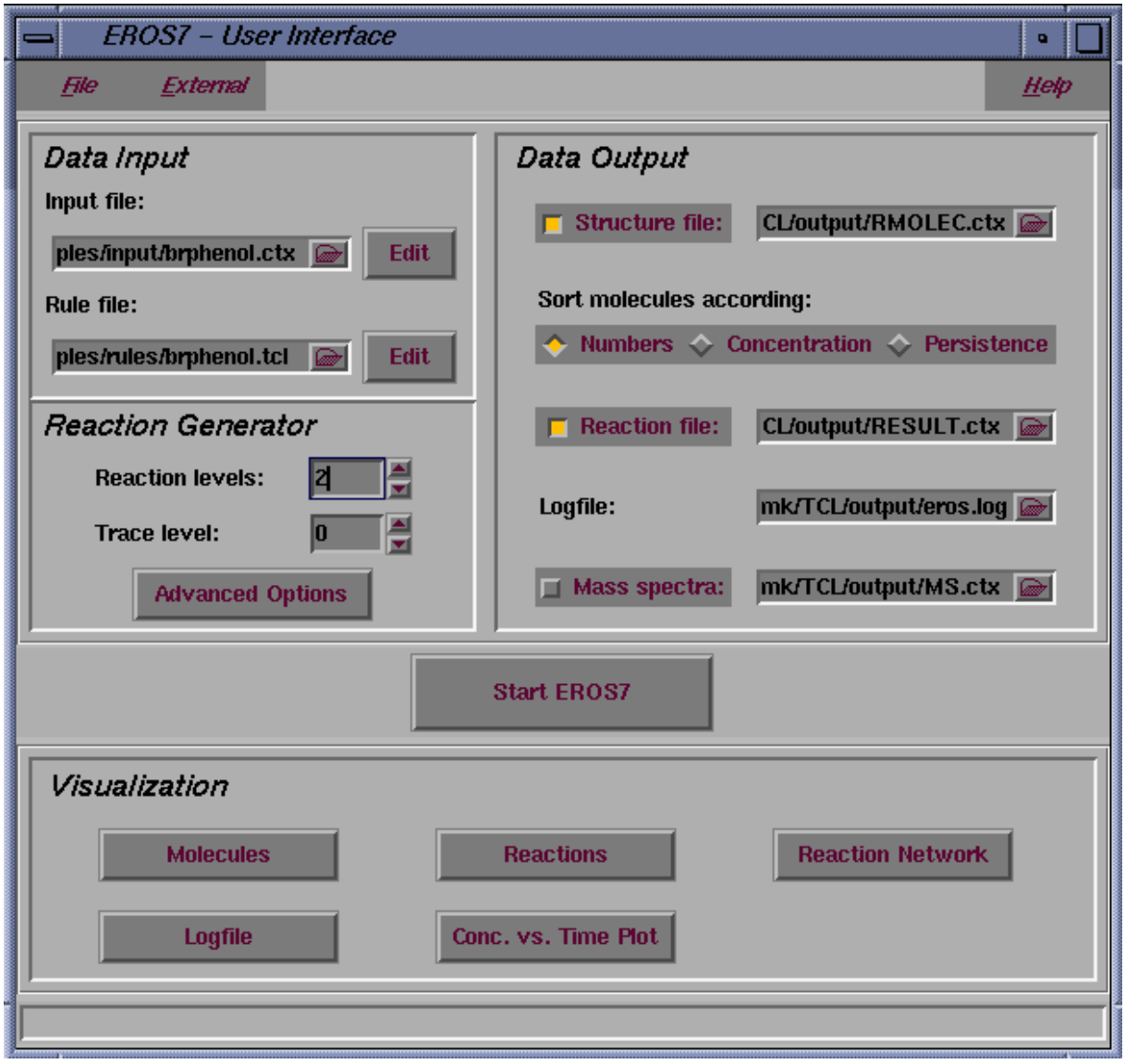


Figure 2-1. The main command center.

In general, all interactions can be performed by clicking buttons with the left mouse button or via keyboard strokes. Pressing the "tab" keys alters the activated widget, which can be set to action by pressing "enter". In the following chapters a description of the capabilities of the graphical user interface will be given.

Guided tour: As you read along, you are invited to perform your first EROS7 run following the instructions given in italics at the end of each chapter. This will allow you to simulate the degradation of s-triazines in soil to illustrate the capabilities of the EROS system and the CACTVS tools.

2.2 Input Specifications

General note:

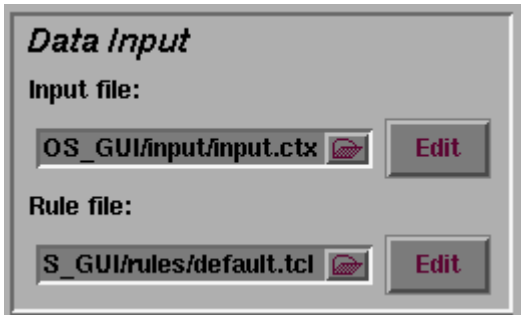


Figure 2-2. Data input.

2.2.1 Structure Input

The input of the structure of chemical compounds can be made in two ways. If the structures are already available in the CTX file format, you can just click on the tiny folder symbol next to the default file name to open a file browser and select the appropriate file. All sample input files are stored in the directory input relative to the directory you chose to copy the sample files when you started the graphical user interface the first time (usually /home/user/eros7/input).
This file may contain one or more chemical structures serving as starting materials for reaction prediction. The other possibility is to start the CACVTS Editor by clicking the "edit" button to create or edit a structure file. For detailed information on the editor please refer to the CACTVS-Editor manual (see Appendix 7.1 or http://www2.ccc.uni-erlangen.de/cactvs/index.html). Other structure editors can easily be appended to the EROS system as communication is performed through the CTX ASCII file format and conversion programs to standard structure exchange formats such as MDL SDFile, SMILES strings etc. are available.

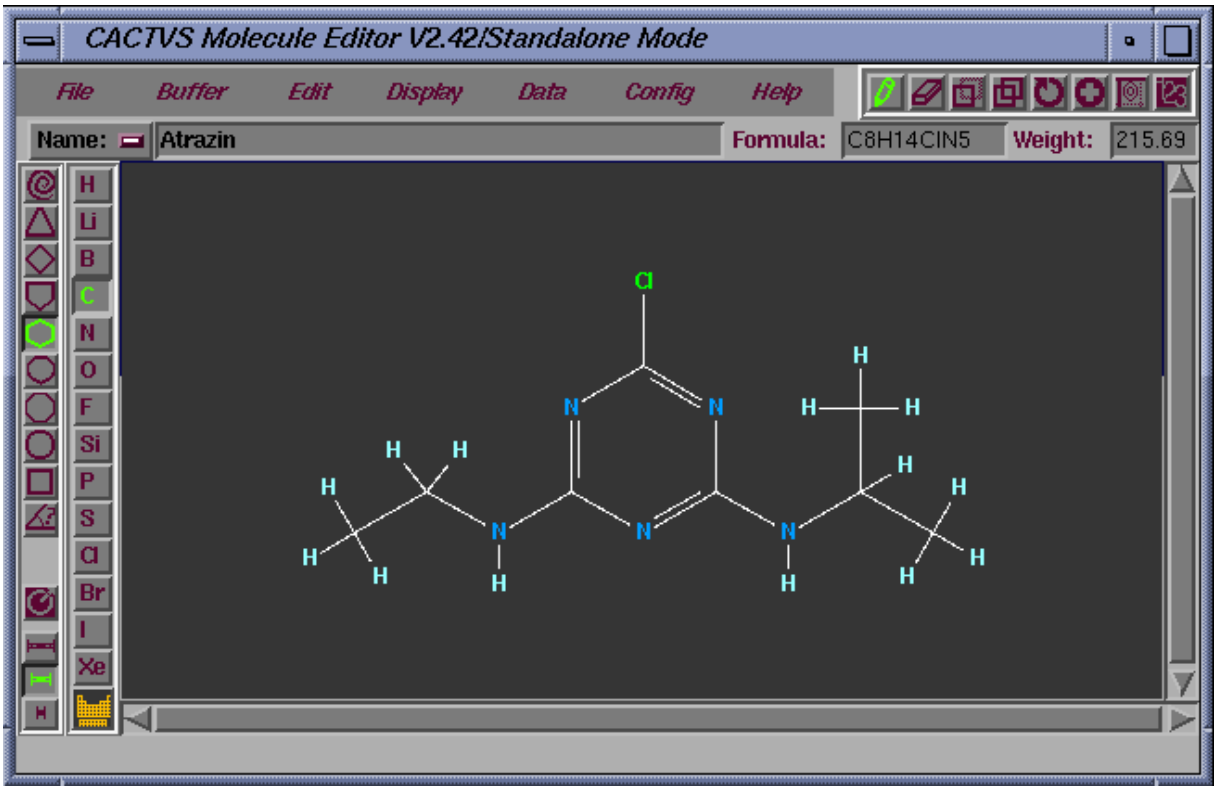


Figure 2-3. The CACTVS Editor.

Guided tour: Please select the file named "atrazine.ctx" in your input directory. Alternatively draw the structure of atrazine using the CACTVS Editor.

2.2.2 Selecting a Rule File

The chemical knowledge of EROS7 is stored in rule files in ASCII format. Select an appropriate rule file by clicking the folder symbol next to the default file name. A detailed description of existing reaction rules can be found in the section "Sample EROS Runs". If you are familiar with the scripting language *Tcl*, you can also edit existing rule files for your needs, which are stored in the directory rules relative to the directory you selected during the first start of the EROS GUI for the samples (usually /home/user/eros7/rules). The "edit" button will start a text editor with the displayed file name. For details about EROS7 reaction rules please refer to the chapter 4 "Writing Your Own Reaction Rules".

Guided tour: Select the rule file named "triazine.tcl" in your "rules" directory. In this file the main degradation pathways of s-triazines in soil, hydrolysis and reductive dealkylation, are coded for the use with EROS7.

2.3 Reaction Generator

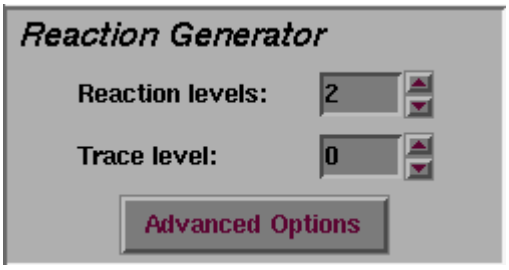


Figure 2-4. Setting options for the reaction generator.

In this section all command line options concerning the reaction generators can be set.

Reaction levels:
Increase or decrease the number of desired reaction levels by clicking the up and down arrows next to the number. Alternatively you can click on the number and type in any number higher than zero.

Trace output:
Increase or decrease the amount of trace output written to the logfile. Choose zero for minimum output up to 5 for maximum trace output.

Advanced Options:
By clicking the button "advanced options" a configuration window will appear. Usually, you will not need this options, but in some special cases, e.g., debugging reaction rules, they might be considered as useful.

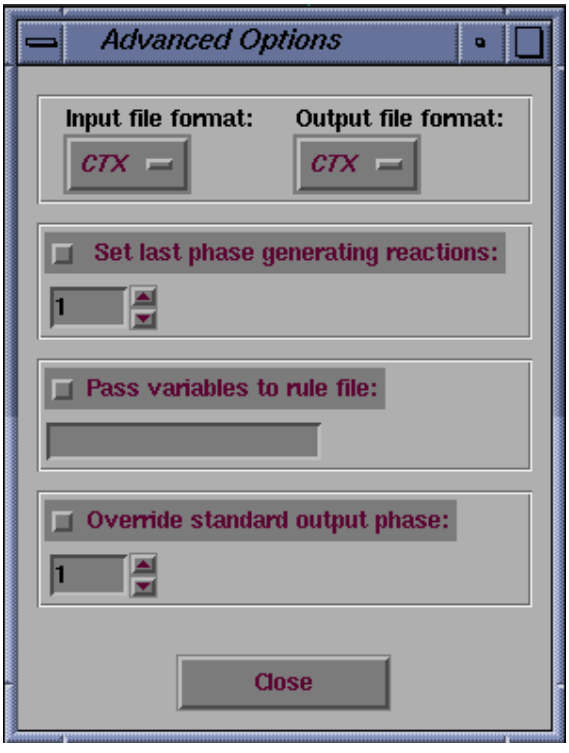


Figure 2-5. Configuration of advanced options.

Usually, the input of structures will be given as connection tables in the CTX format. Internally, this valence bond representation is converted into a representation that is more related to a MO description (MOSES) so that reaction prediction can deal with the full scope of organic chemistry. For output, the structures are reconverted to a connection table representation in the CTX format. This behavior can be changed by choosing the MOSES format either as input or output format.
The phases in a rule file are numbered from 1 to n. If a reaction rule has 5 phases, all generating reactions, it is possible to stop reaction generation after phase three by checking the box and setting the counter to three.
The next checkbox is used to create and transfer additional variables to the rule file. The variables have to be separated by a comma and follow the syntax: name-separator-value (no blanks at all!). The symbol of the separator also indicates the kind of the variable created: "#" for an integer, ":" for a double and "$" (you have to type \$ for the $ because of the shell wrapper) for const char *, (e.g. pi:3.14,num#7).
The structure file written by EROS7 represents the molecules of that phase that was defined as the "output_phase" in the rule file. With the last checkbox active, it is possible to redirect the molecules of any other phase to the structure file.
Guided tour: For our example please set "reaction levels" to 9 to simulate the degradation all the way to the formation of carbon dioxide, ammonia, and water. Setting trace output or advanced options is not necessary in our example.

2.4 Data Output

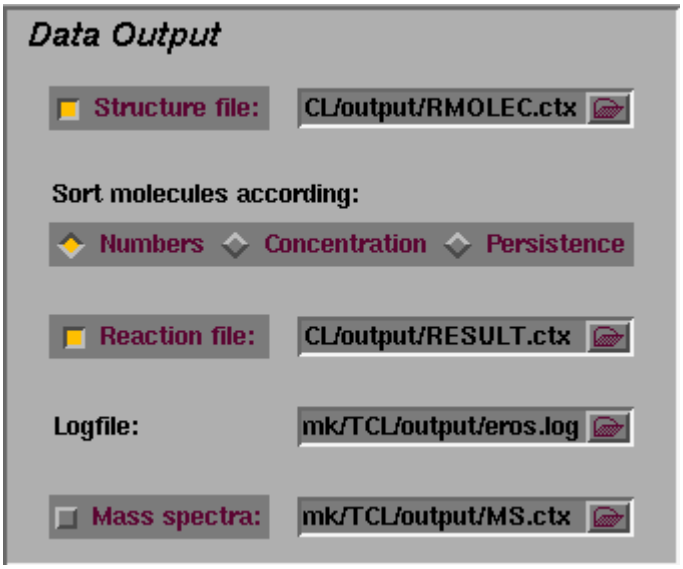


Figure 2-6. Configuration of data output.

In this section the output of a program run can be configured. In most cases, the structure and reaction file will be sufficient. The structure file lists all different molecules handled in this program run. The molecules can be sorted by molecule numbers (default), by concentration or by their persistence. The reaction file stores the molecules for every generated reaction. Change the default file names according to your needs simply by editing the file entries or clicking the folder symbols to browse your directories for existing file names. If any of the file checkboxes are unchecked the respective file will not be written (the EROS7 logfile will be written in any case).

If you are using EROS7 reaction rules designed for the simulation of mass spectra, check the "Mass spectra" button to write the generated mass spectra to a file.

Guided tour: Make sure that the buttons "Structure file" and "Reaction file" are checked. Sort the molecules in the structure file by their persistence.

2.5 Visualization of Results

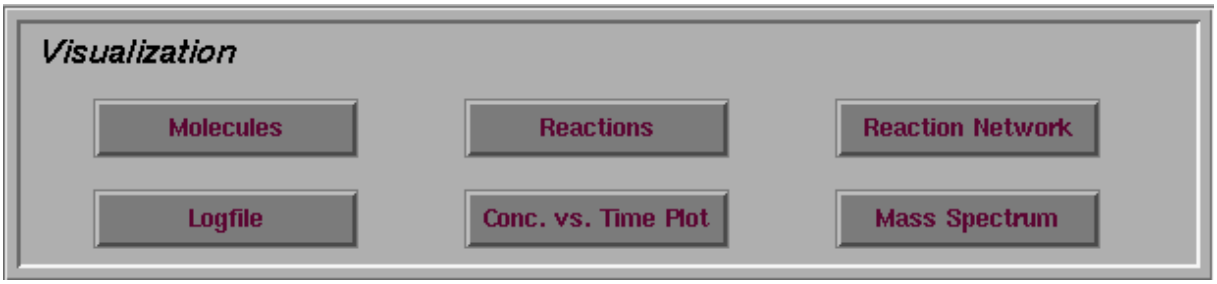


Figure 2-7. Visualize the EROS7 results.

After the first EROS7 program run a "Visualization" frame will appear. These buttons are generated according to the selected "Data Output" options. This means if the buttons for "Structure file" and "Mass spectra" were unchecked, the visualization of the structure file and the mass spectra is not possible.

Molecules:
This button will open the CACTVS Browser *csbr* to display the molecules of the EROS7 structure file.

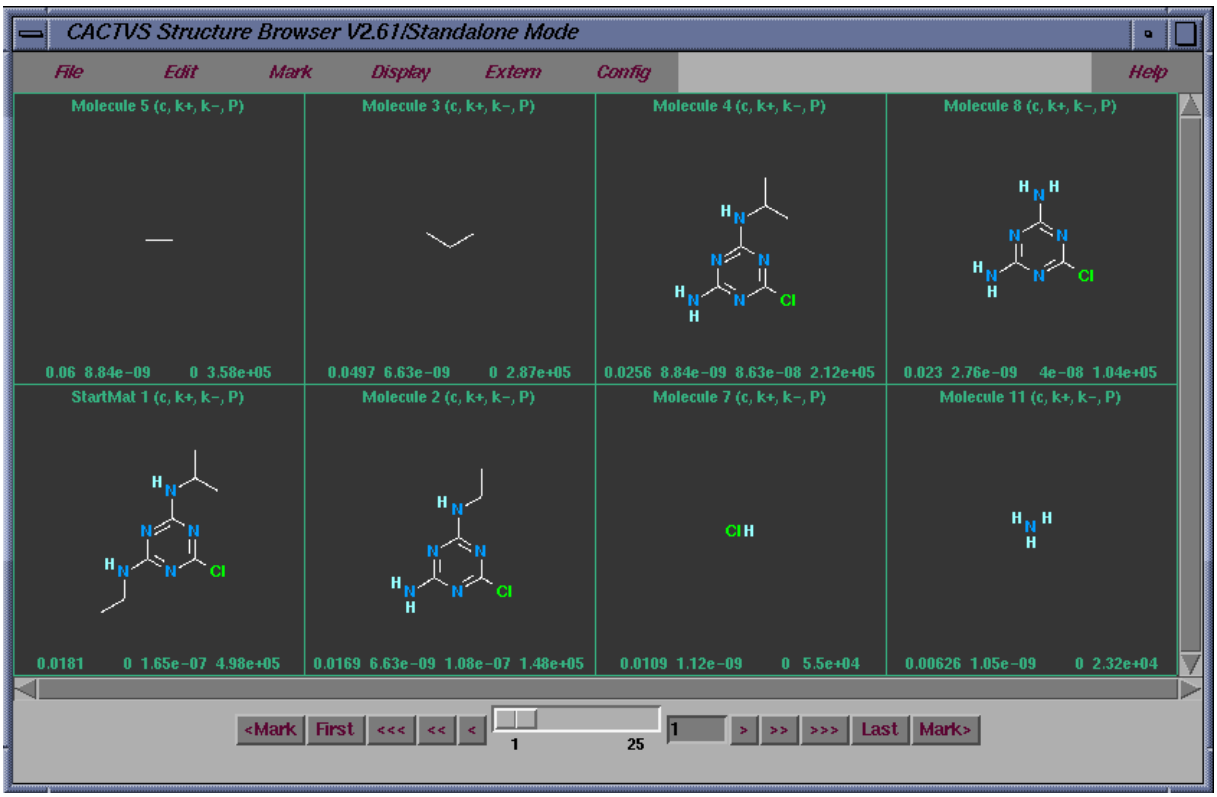


Figure 2-8. Visualization of the EROS7 structure file with the CACTVS Browser.

Guided tour: Please click on "Molecules" to visualize the EROS7 structure file. For each molecule, information about the predicted end concentration (c), the overall first order formation and degradation rates (k+, k-) and the persistence of the compound will be displayed. The molecules will be sorted by their persistence (P).

Note: In the case of the appearance of bimolecular reactions and the probability kinetics the values displayed for k+ and k- are not valid.

Reactions:
This will open the CACTVS Browser *csbr* with the EROS7 reaction file. The starting materials and products for each reaction will be displayed connected by an arrow.

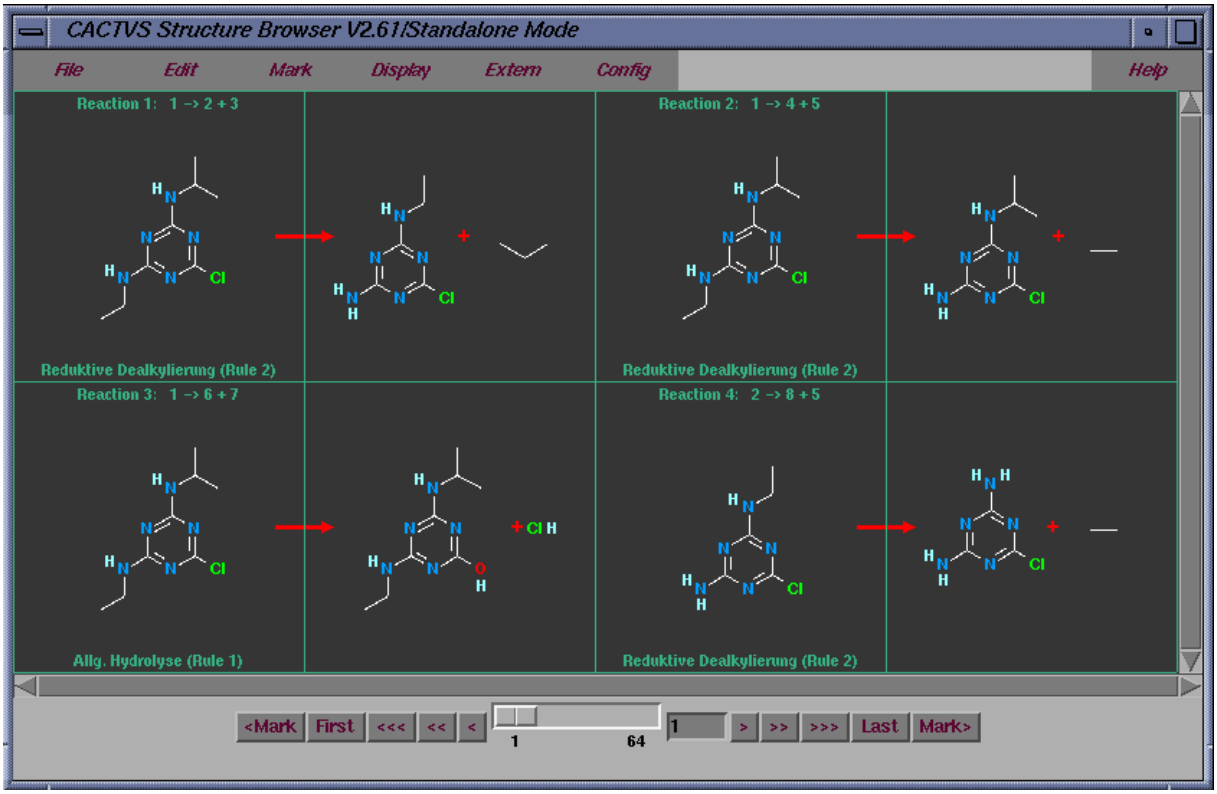


Figure 2-9. Visualization of the EROS7 reaction file with the CACTVS Browser.

Guided tour: Click the "Reactions" button to see starting materials and products for each reaction.

Reaction network:
By clicking this button the EROS7 reaction file will be visualized as a reaction tree by CACTVS Tree *cstr*. For details about CACTVS Tree please refer to the manual (see Appendix 7.2 or http://www2.ccc.uni-erlangen.de/Oliver_Sacher/cstree/frame.html). CACTVS Tree was designed to display only reactions that were generated by EROS7 with the reaction mode "MONOMOLEC" set in the rule file

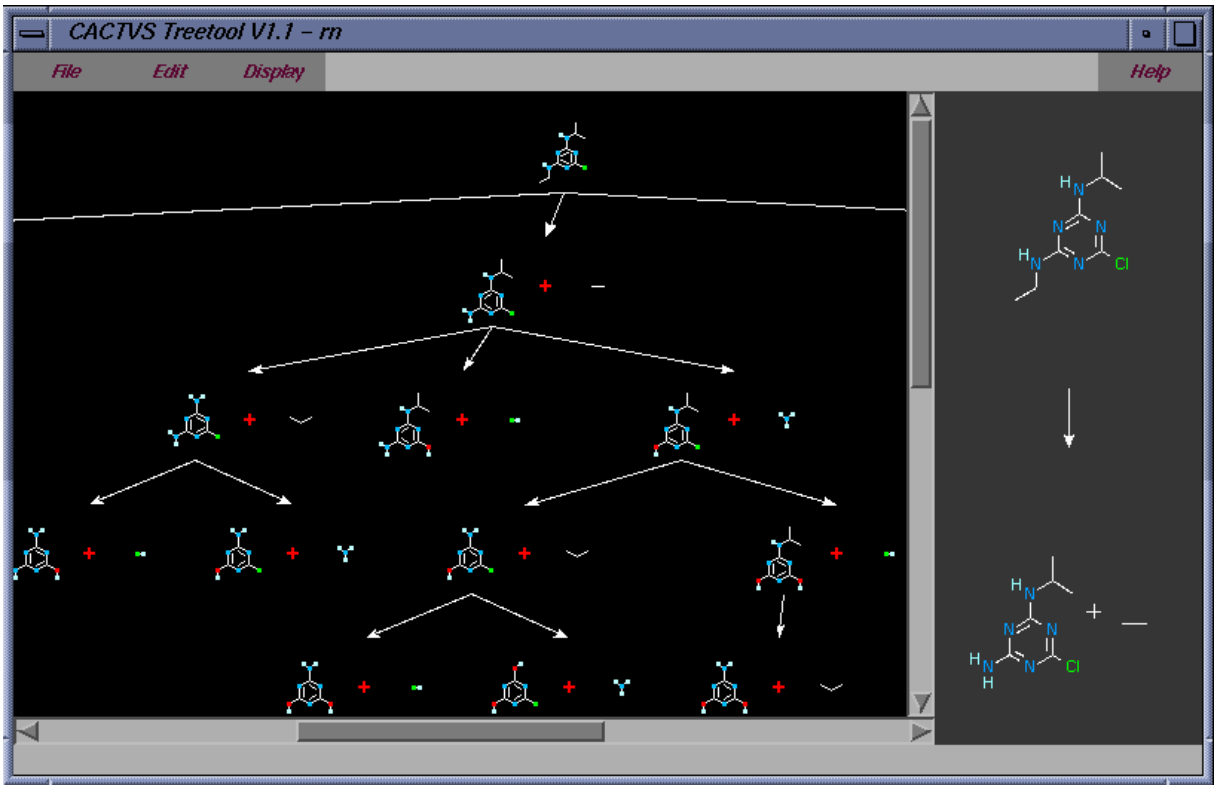


Figure 2-10. Visualization of the EROS7 reaction file with CACTVS Tree.

Guided tour: Visualize the reaction tree by clicking "Reaction Network".

Logfile:
Clicking this button will start the default text editor with the EROS7 logfile.
Guided tour: If you like, view the EROS7 logfile by clicking the appropriate button.

Concentration Time Plots
Click this button to see the concentration time plots of all compounds involved in this EROS run. This feature is enabled when kinetic calculations were done with Gear's algorithm (to be set in the rule file). The plots will be displayed for 120 seconds. Additionally a postscript file of the curves will be written to the output directory. If there is no valid data an empty plot with an error message will appear.

Mass Spectra:
Simulated mass spectra will be displayed with the spectra visualization tool CACTVS IR *csir*.

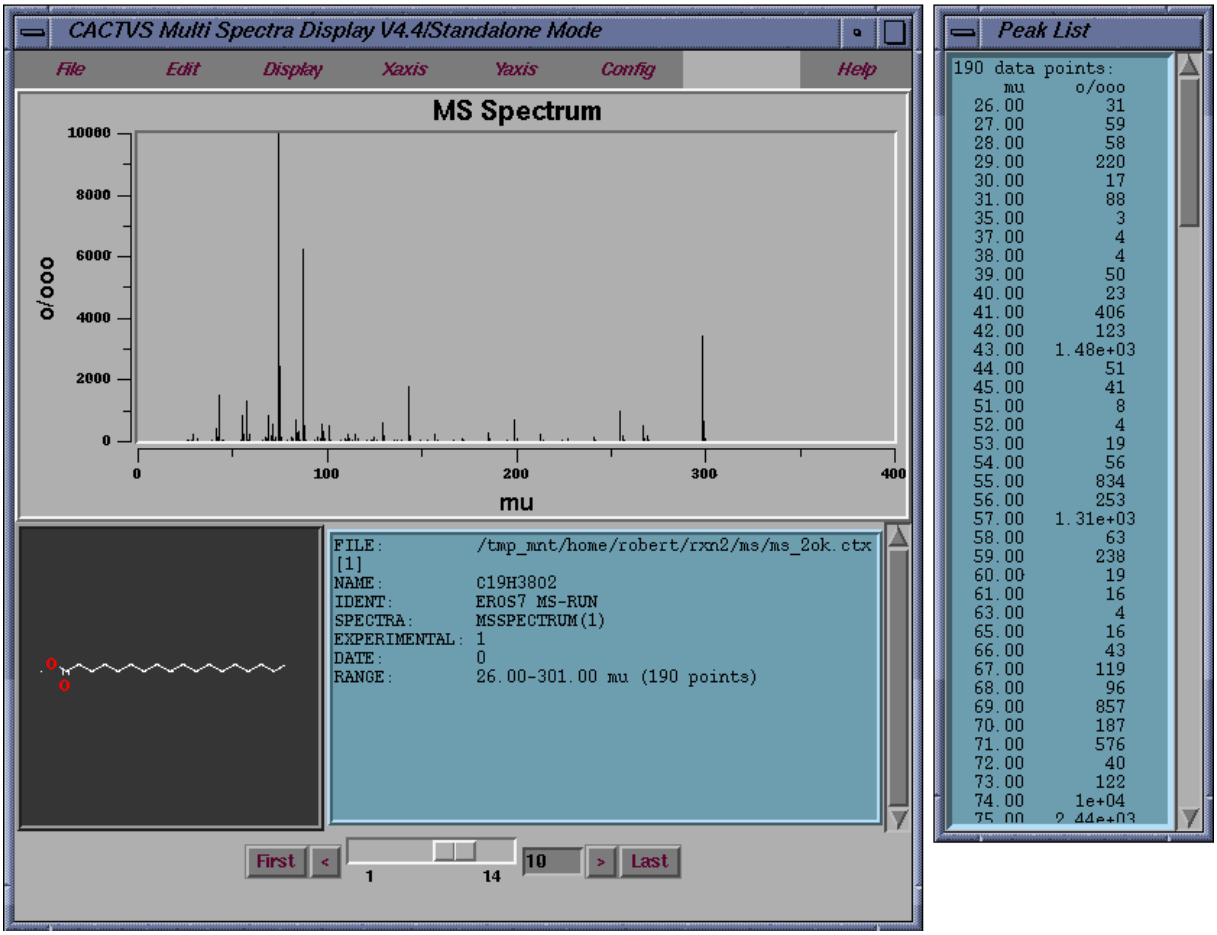


Figure 2-11. Visualization of predicted mass spectra with CACTVS IR.

3 Sample EROS Runs

In order to show the user the range of applications that can be made with EROS and to make her/him more familiar with running EROS, a series of sample runs are provided.
These studies include

- an introduction into the problem - a discussion of the essential elements of the reaction rules and how they are derived
- the input specifications
- the corresponding reaction rules
- the input file for running EROS

The examples start with a simple reaction without any reactivity evaluation (3.1), then cover combinatorial chemistry experiments, again without any evaluation of the reactions (3.2). A further example illustrates how EROS can be used for the exhaustive generation of isomers, (3.3).
Then, reactions are presented that include an evaluation of chemical reactivity. First, a simple laboratory reaction is presented (3.4). Then, the degradation of chemicals in the environment is dealt with (3.5). Next, the breakdown of a molecule in the mass spectrometer is modeled to simulate a mass spectrum (3.6).
The final example shows the combination of two reaction environments, a laboratory synthesis followed by the mass spectroscopy of the starting materials and reaction products (3.7).

3.1 The Synthesis of an Amide

Essentials to learn:

- how the EROS system works
- reactions as bond rearrangements
- the use of phases

Description:

This example shows how the EROS system can model a simple reaction like the synthesis of an amide. In this example, acetic acid reacts with methylamine to give an amide. Figure 3-1 shows which bonds are broken and which ones are made in this reaction.

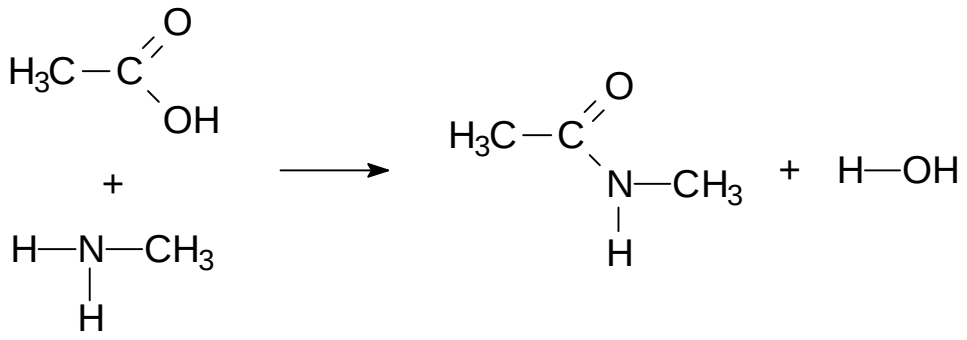


Figure 3-1. Amide synthesis from acetic acid and methylamine.

Clearly, the reaction proper does not follow this pathway, as an acid and an amine gives a salt in a proton transfer reaction. The reaction shown in Figure 3-1 only models the net result of an amide synthesis from an acid and an amine which has to be initiated by activation of the acid as an acid chloride or an ester. However, as no evaluation of chemical reactivity is attempted in this example, such a treatment of the overall changes in a sequence of reaction steps is allowed. First, all the starting materials are put into phase 1. Then, the constraints in RULE_21 check if one of the chemicals is an acid. If there is an acid, it is moved into phase 2. In the same way, an amine is moved into phase 3.

Then, RULE_32 makes the rearrangement of atoms and bonds and transfers the products into phase 4 (mode: INERT). Phase 3 (mode: SURFACE) is in contact with phase 2 (mode: INERT) and generates the reaction by taking a molecule from phase 2 and a molecule from phase 3.

Finally, an output file is generated.

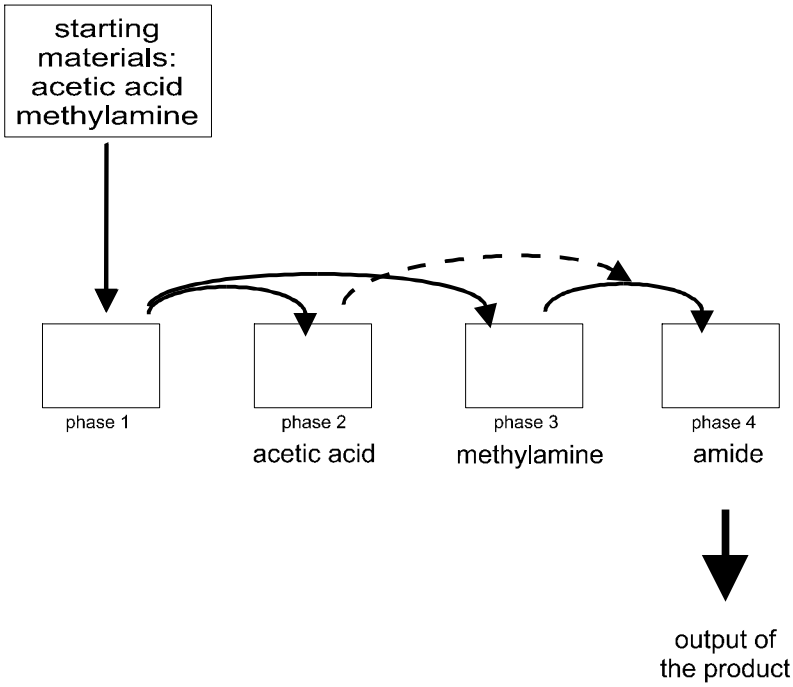


Figure 3-2. Scheme of phases needed in the reaction of acetic acid with methylamine.

Main features of the reaction rule file:

Rule header:

- number of reactors: 1
- number of phases: 4
- mode of phase 1: MONOMOLEC
- mode of phase 2: INERT
- mode of phase 3: SURFACE
- mode of phase 4: INERT

Reaction rules:

Reaction type 1 in RULE_21:

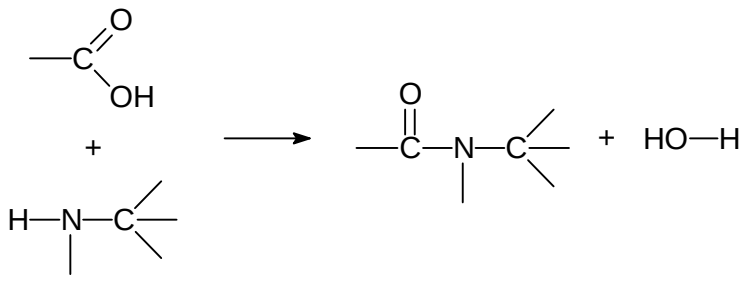
Check for acid group and transport acids into phase 2

Reaction type 2 in RULE_23:

Check for amino group and transport amines into phase 3

Reaction type 3 in RULE_32:

Amide formation



Reaction Generator:

Reaction levels: 1

Trace level: 0

Reaction rule file and input file:

The reaction rule and the CTX-input are contained as files (amide.tcl and amide1.ctx) on the CD-ROM for the distribution of the EROS system.

3.2 Combinatorial Chemistry

In this section, the use of EROS for modeling Combinatorial Chemistry experiments is explained.
In Combinatorial Chemistry two or more sets of molecules react with each other. Each molecule of the first group reacts with each molecule of the second group to give in all possible combinations of reactions all conceivable combinations of products. This can be achieved in a variety of experimental set-ups such as in parallel synthesis or liquid phase experiments.

3.2.1 The Synthesis of Esters and Amides

Essentials to learn:

- how to set up the modeling of combinatorial chemistry experiments
- the concept of phases in parallel synthesis
- a reaction rule that can be applied both to the synthesis of amides and esters

Description:
The first example is the combinatorial synthesis of esters and amides.
The following reactions can be handled

- acids or acid chlorides and amines react to amides
- acids or acid chlorides and alcohols react to esters

First, the starting materials are put into phase 1 that has the mode MONOMOLEC.
From this phase, acids and acid chlorides are moved into phase 2 (mode: INERT). This is caused by the constraints in RULE_21 and RULE_22 in the reaction rule file. The constraints check if the molecule has an acid group or an acid chloride group. The amines and alcohols are moved into phase 3 (RULE_23 and RULE_24 in the reaction rule file; the constraints check for the presence of an NH_2 - group or an OH - group). The phase 3 has the mode SURFACE and the phase 4 the mode INERT (Figure 3-3).

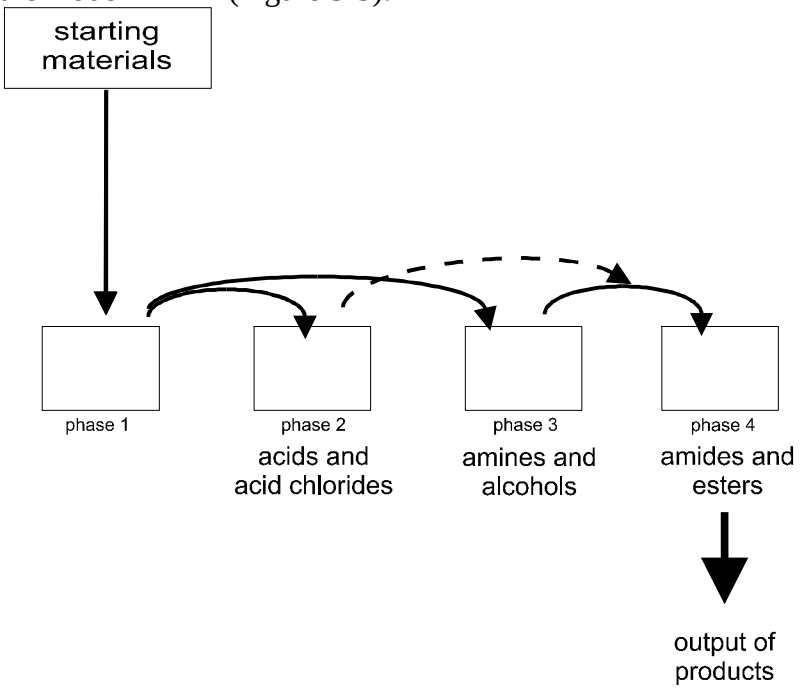


Figure 3-3. Combinatorial synthesis of amides and esters.

The amines and alcohols from phase 3 (mode: SURFACE) react with the acids and acid chlorides of phase 2 (mode: INERT) that is in contact with phase 3 (RULE_32 in the reaction rule file). The constraints in this rule search for the reaction substructures that are needed to give amides and esters. Then a simple rearrangement of atoms and bonds take place (Figure 3-4). The products arrive in phase 4 that has the mode INERT. Phase 4 is the output phase. The content of the reaction output file are all reactions made in this EROS run, the moves of the starting materials from phase 1 to phase 2 and 3, too.

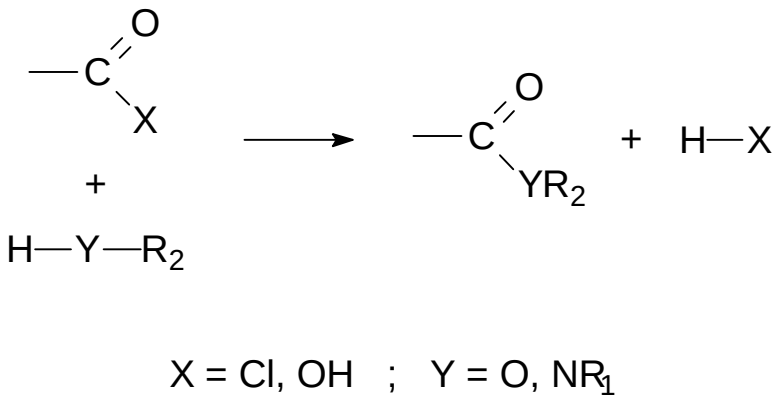


Figure 3-4. Bond rearrangement and constraints on the atoms of the reaction type used.

The output file for the structures comprises all products from the output phase 4.
The two CTX output files can be watched with the CACTVS browser *csbr*.

Main features of the reaction rule file:

Rule header:

- number of reactors: 1
- number of phases: 4
- mode of phase 1: MONOMOLEC
- mode of phase 2: INERT
- mode of phase 3: SURFACE
- mode of phase 4: INERT

(see chapter 1.3.1.3 and 3.1)

Reaction rules:

Reaction type 1 in RULE_21:
Check for acid group and transport acids into phase 2

Reaction type 2 in RULE_22:
Check for acid chloride group and transport acid chlorides into phase 2

Reaction type 3 in RULE_23:
Check for amine group and transport amines into phase 3

Reaction type 4 in RULE_24:
Check for hydroxyl group and transport alcohols into phase 3

Reaction type 5 in RULE_32:
Condensation (see Figure 3-4)

Reaction Generator:
Reaction levels: 1
Trace level: 0

Reaction rule file and input file:
The reaction rule and the CTX-input are contained as files (amide.tcl and amide2.ctx) on the CD-ROM for the distribution of the EROS system.

3.2.2 The Synthesis of Pyrazoles

Essentials to learn:
- How to code extensive bond rearrangements

Description:
In combinatorial chemistry different sets of starting materials are combined in all conceivable variations to synthesize a wide range of compounds.
Pyrazoles can be obtained from 1,3-diketones and substituted hydrazines.

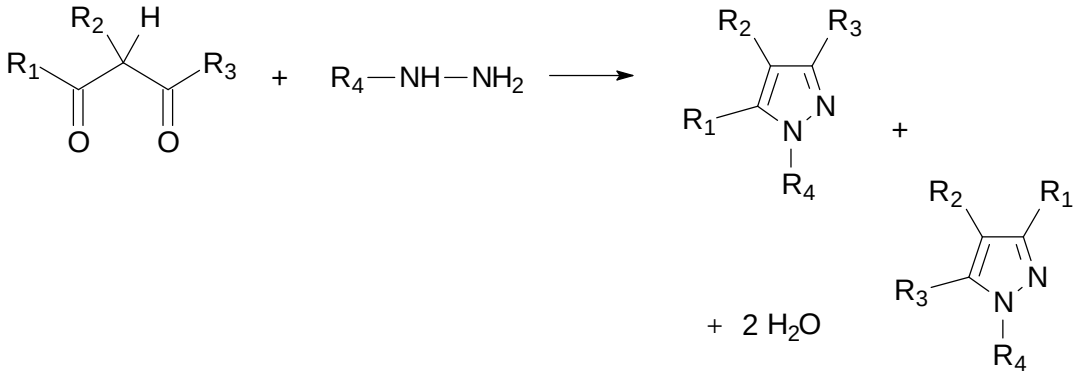


Figure 3-5. Synthesis of pyrazoles.

In terms of number of bonds broken and made, this reaction involves quite an extensive electron rearrangement. Such extensive reaction schemes can also be coded in a reaction rule.
In a combinatorial chemistry experiment two or more sets of starting materials react in such a way as to react each molecule from set one with each molecule of set two, etc. In our example, a set of two 1,3-diketones will be reacted with three substituted hydrazines.
The two sets of starting materials are shown in Figure 3-6.

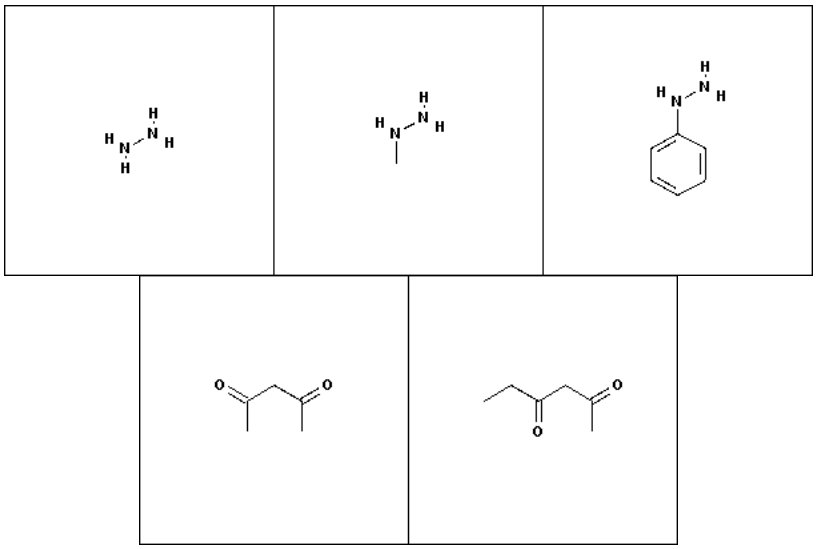


Figure 3-6. Starting materials of the combinatorial synthesis of pyrazoles.

After input of these five molecules, they are assigned to two different phases depending on the substructure they contain: The 1,3-diketones are assigned to phase 2 whereas the hydrazines are put into phase 3. Then, from each phase one molecule is taken out one at a time and reacted with a molecule of the other phase. As unsymmetrical 1,3-diketones can give rise to two different pyrazoles, (see Figure 3-5) both ways of combining 1,3-diketones with a hydrazine are explored. The products are stored in phase 4 (Figure 3-7).

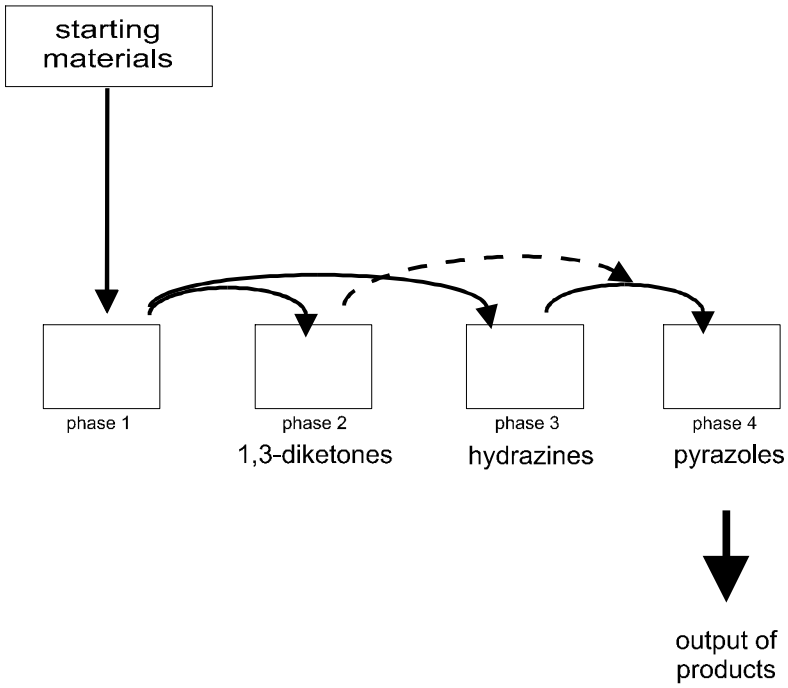


Figure 3-7. Scheme of phases for the synthesis of pyrazoles.

The results of this program run are shown in Figure 3-8.

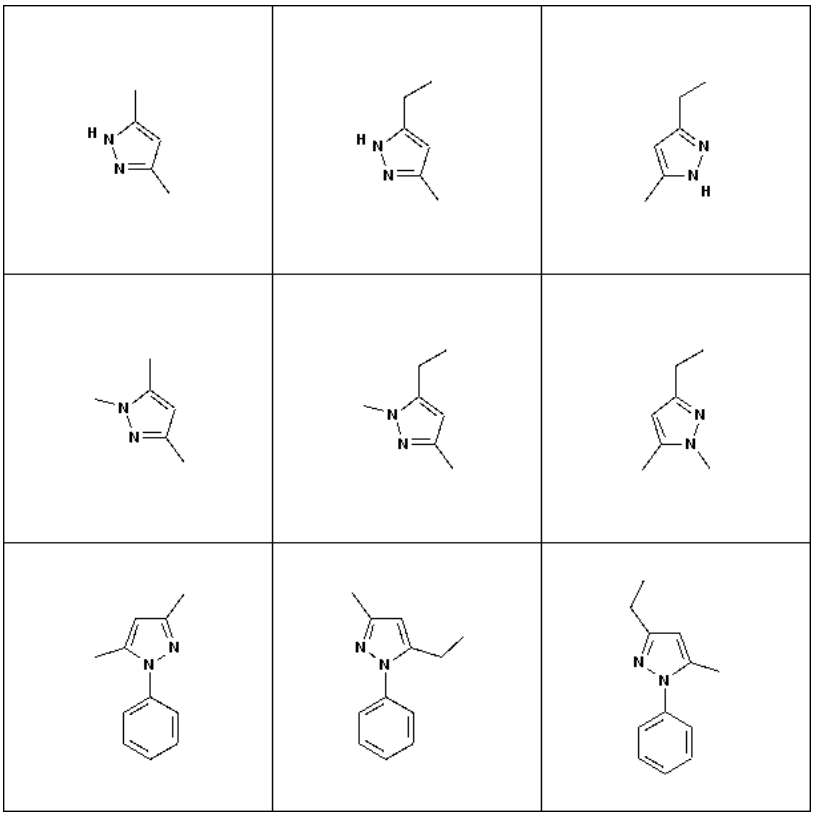


Figure 3-8. Products of the combinatorial synthesis of pyrazoles.

Main features of the reaction rule file:

Rule header:

- number of reactors: 1
- number of phases: 4
- mode of phase 1: MONOMOLEC
- mode of phase 2: INERT
- mode of phase 3: SURFACE
- mode of phase 4: INERT

(see chapter 1.3.1.3 and 3.1)

Reaction rules:

Reaction type 1 in RULE_11:

Check for the substructure of 1,3-diketones (see Figure 3-5) and transport 1,3-diketones into phase 2

Reaction type 2 in RULE_12:

Check for the substructure of hydrazines (see Figure 3-5) and transport hydrazines into phase 3

Reaction type 3 in RULE_13:

Pyrazole formation (see Figure 3-5)

Reaction Generator:

Reaction levels: 1

Trace level: 0

Reaction rule file and input file:
The reaction rule and the CTX-input are contained as files (pyrazole.tcl and pyrazole.ctx) on the CD-ROM for the distribution of the EROS system.

3.2.3 1,4-Benzodiazepines

Essentials to learn:

- How to set up the modeling of multistep parallel synthesis

Description:
Whereas the previous example of the synthesis of pyrazoles deals with a one-step reaction, the example given here deals with a two-step synthesis. In particular, an experiment by Ellman and coworkers [24] for the synthesis of 1,4-benzodiazepines (see Figure 3-9) will be analyzed.

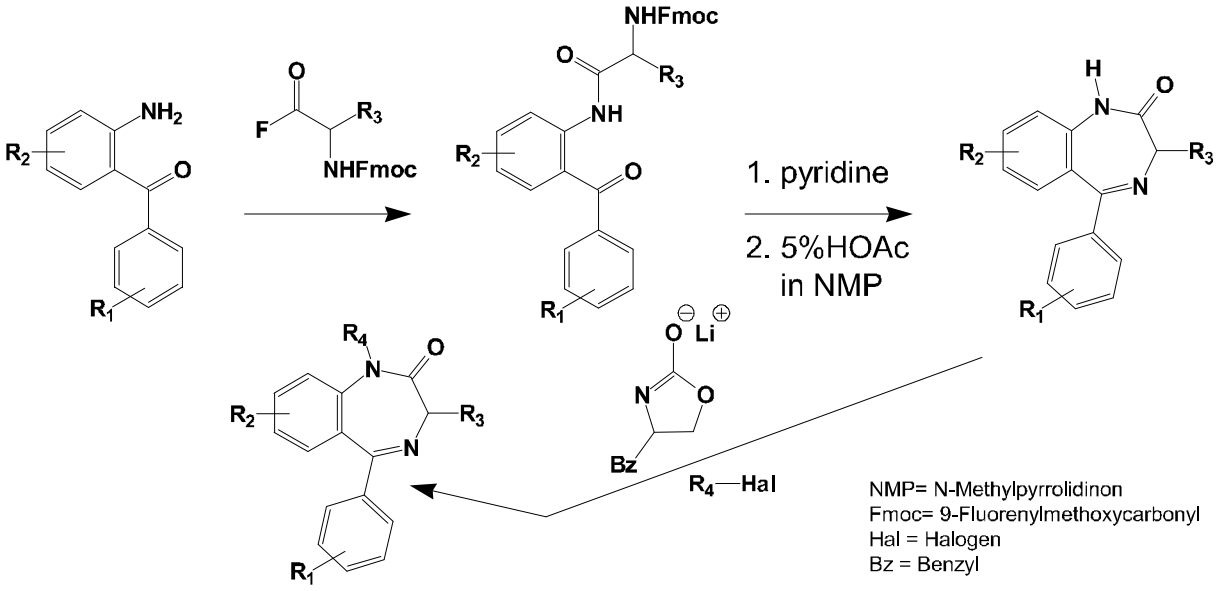
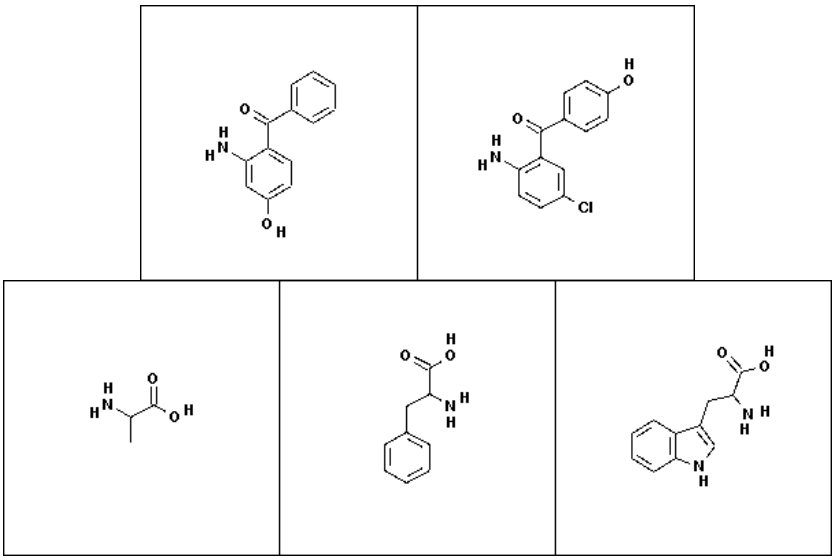


Figure 3-9. Synthesis of 1,4-benzodiazepines by Ellman and coworkers.[24]

Three sets of starting materials are necessary to perform this sequence, 2-aminobenzophenones, amino acids, and alkylating agents. For modeling this experiment with EROS, the two 2-aminobenzophenones, three amino acids, and three alkylating agents shown in Figure 3-10 were chosen.



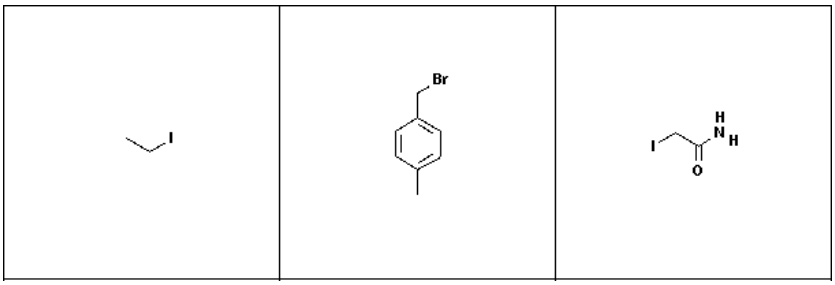


Figure 3-10. Starting materials of the combinatorial synthesis of 1,4-benzodiazepines.

The entire reaction sequence is modeled by six phases (see Figure 3-11), one phase for storing all starting materials, three phases for storing the three different sets of these starting materials, and two phases for performing the two major reaction steps, the combination of the 2-aminobenzophenone with an amino acid followed by cyclisation to 1,4-benzodiazepines, and the alkylation step at nitrogen-1 of this ring system.

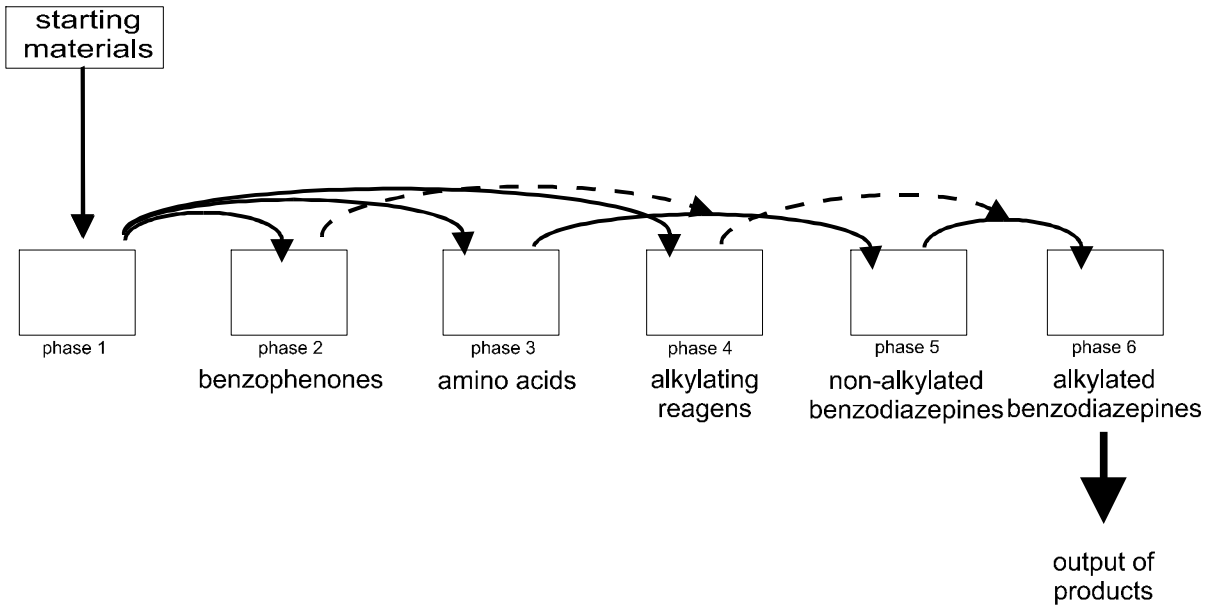


Figure 3-11. Scheme of phases for the synthesis of 1,4-benzodiazepines.

Figure 3-12 shows the structures obtained in this combinatorial chemistry experiment. All 18 conceivable structures (2 x 3 x 3) were obtained in this run.

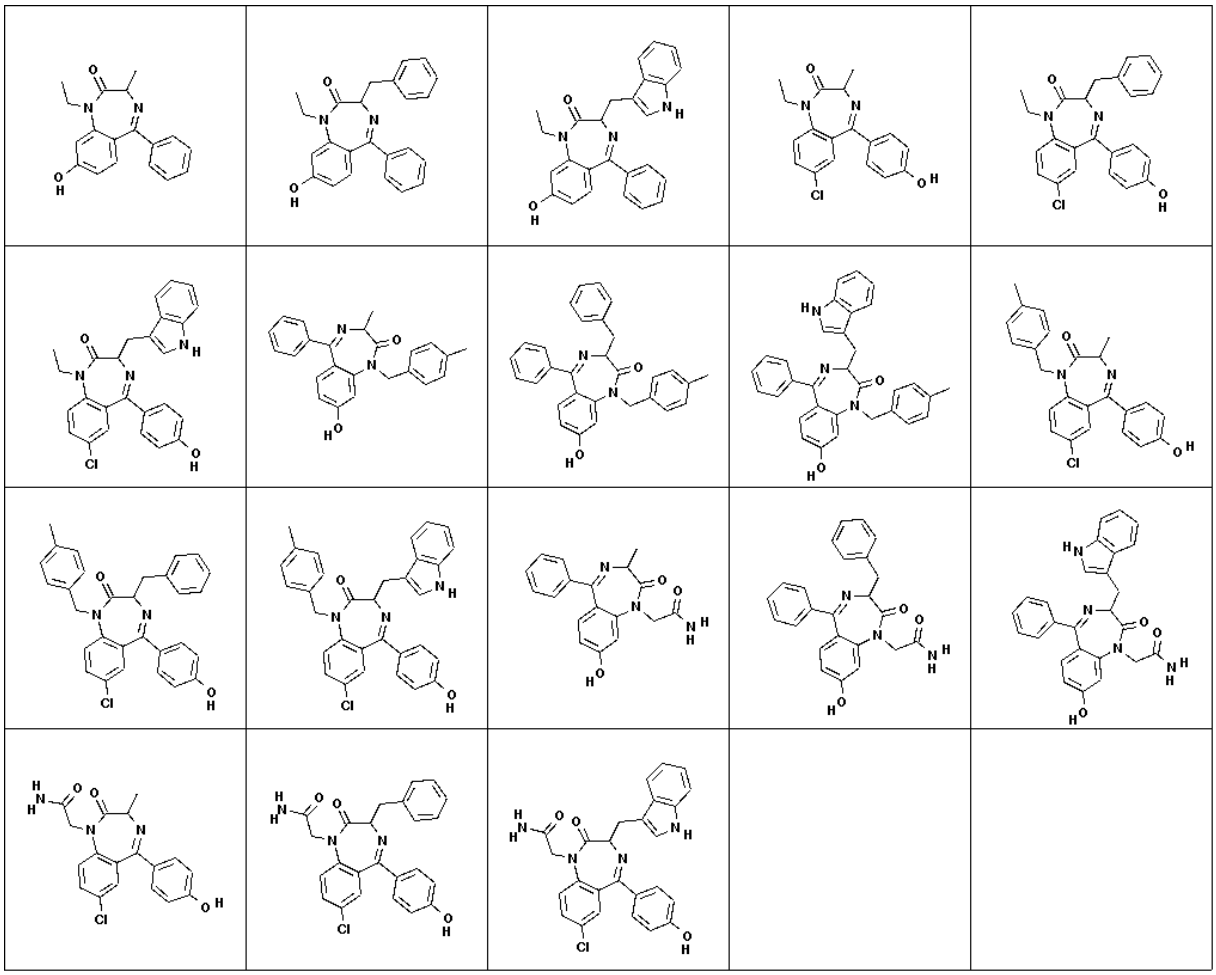


Figure 3-12. Products of the combinatorial synthesis of 1,4-benzodiazepines.

Main features of the reaction rule file:

Rule header:

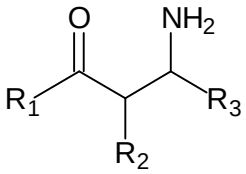
- number of reactors: 1
- number of phases: 6
- mode of phase 1: MONOMOLEC
- mode of phase 2: INERT
- mode of phase 3: SURFACE
- mode of phase 4: INERT
- mode of phase 5: SURFACE
- mode of phase 6: INERT

(see chapter 1.3.1.3)

Reaction rules:

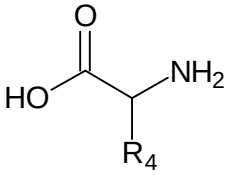
Reaction type 1 in RULE_21:

Check for the substructure and transport 2-aminobenzophenones into phase 2:



Reaction type 2 in RULE_22:

Check for the substructure and transport amino acids into phase 3:



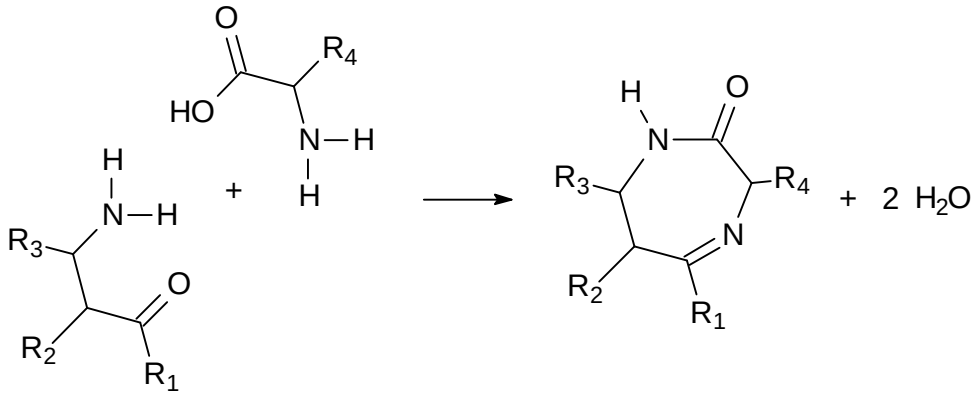
Attention: Don´t use asparagine (Asn) or glutamine (Glu) as amino acids, because the alkylating agents react with the free amide groups of Asn and Glu in the non-alkylated benzodiazepines, too.

Reaction type 3 in RULE_23:

Check for a chlorine, bromine or iodine atom with a bond to an aliphatic carbon atom and transport of alkylating agents into phase 4

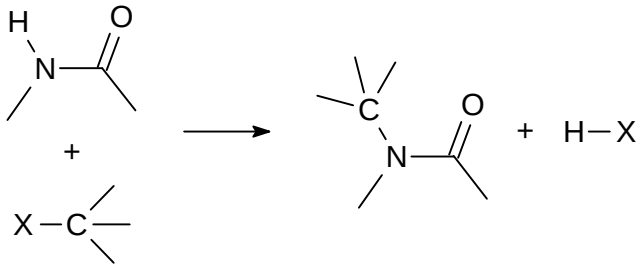
Reaction type 4 in RULE_31:

Ring closure and formation of an unsubstituted 1,4-benzodiazepine:



Reaction type 5 in RULE_32:

Alkylating reaction:



Reaction Generator:
Reaction levels: 2
Trace level: 0

Reaction rule file and input file:
The reaction rule and the CTX-input are contained as files (diazep.tcl and diazep.ctx) on the CD-ROM for the distribution of the EROS system.

3.3 Exploration of all Potential Reaction Products: Chlorination of Benzodioxine

Essentials to learn:

- Exhaustive exploration of all reaction products
- Suppression of duplicate reaction products

Description:

This example shows how EROS can be applied to the exhaustive generation of isomers for a given problem. This is illustrated here with the generation of all mono-, di-, all the way to octasubstituted chloro-dibenzodioxines.

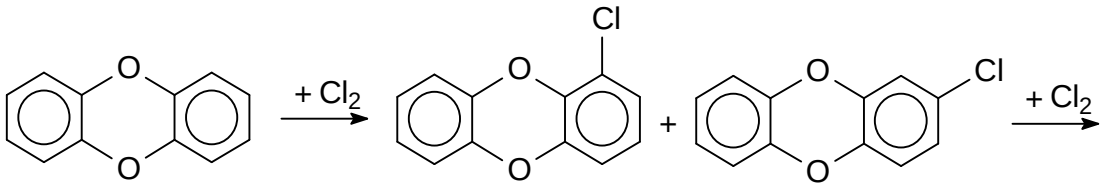


Figure 3-13. Chlorination of benzodioxine.

The first reaction step generates monochlorobenzodioxines, the second dichlorobenzodioxines, etc. Each reaction step is taken care of by an individual phase of the reactor used; the first two phases are needed for the initial storage of the starting materials.

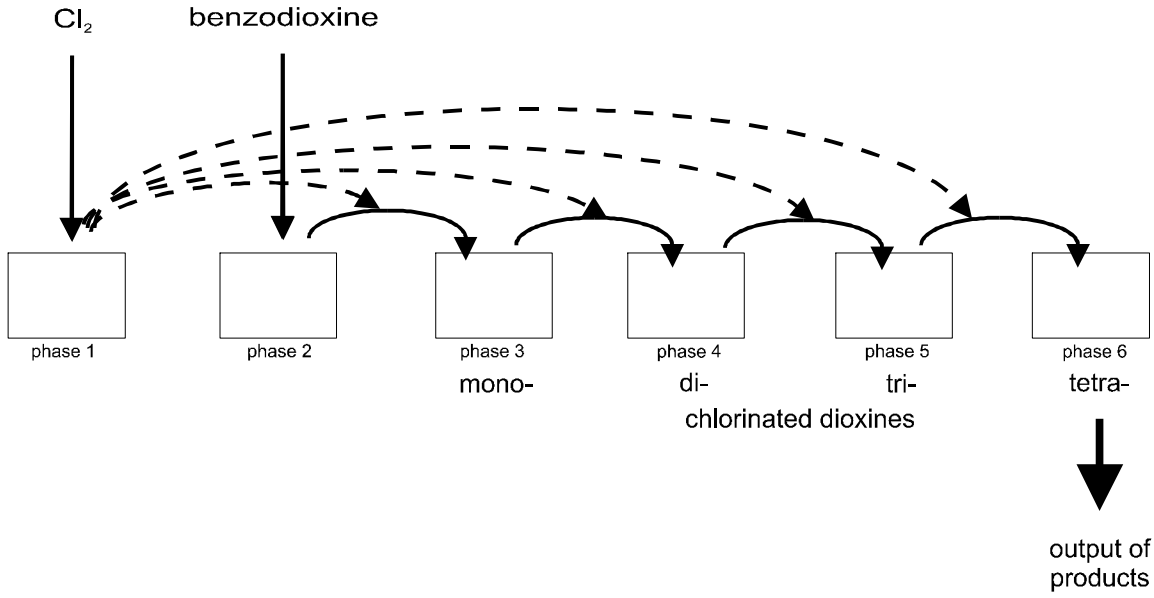


Figure 3-14. Scheme of phases for the chlorination of tetra-chlorinated benzodioxines.

No evaluation of the chlorination of benzodioxine and its substituted derivates is performed as only all possible substituted derivates should be obtained. It is not aspired to estimate the relative ratios of these chlorinated compounds under certain reaction conditions.
The following result is obtained:

2 monochlorobenzodioxines
10 dichlorobenzodioxines
14 trichlorobenzodioxines
22 tetrachlorobenzodioxines

14 pentachlorobenzodioxines
10 hexachlorobenzodioxines
2 heptachlorobenzodioxines
1 octachlorobenzodioxines

This example also shows that the methods for unique identification of a compound, in our case, a hashcode algorithm,[25] work correctly as the one and the same compound can be obtained by different reaction pathways but it will be output only once.

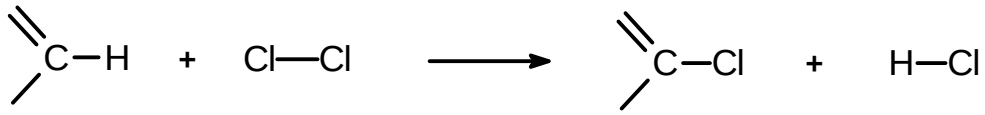
Main features of the reaction rule file:

Rule header:
- number of reactors: 1
- number of phases: 10
- mode of phase 1: INERT
- mode of phase 2 - 9: SURFACE
- mode of phase 10: INERT

(see chapter 1.3.1.3)

Reaction rules:

Reaction type 1:



Reaction Generator:
Reaction levels: 8
Trace level: 0

Advanced Options (optional):
To break off the reaction network transfer an integer variable named phase_inc with the value of 1 to the rule file (see chapter 2.3). The output file contains only not fully chlorinated benzodioxines.
For example: To get all pentachlorobenzodioxines set phase_inc to the value of 1, set the last phase generating reactions to the value of 6, and set the standard output phase to the value of 7 (see Figure 3-14).

Reaction rule file and input file:

The reaction rule and the CTX-input are contained as files (dioxin.tcl and dioxin.ctx) on the CD-ROM for the distribution of the EROS system.

3.4 Multistep Laboratory Reaction: Bromination of Phenol

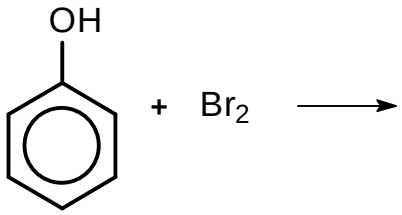


Figure 3-15. Bromination of phenol.

Essential to learn:

- How to derive rules for calculating reaction rates

Description:

This reaction is run in a single vessel, the concentrations of starting materials are such that multiple reactions between the different starting materials might occur. The following specifications are made: one reactor, one phase, reaction mode: MIX.

Next, a reaction rule for the bromination of phenol had to be developed and stored in the knowledge base of the EROS system. The reaction center, i.e., the bonds broken and made in the reaction, was specified as shown in Figure 3-16.

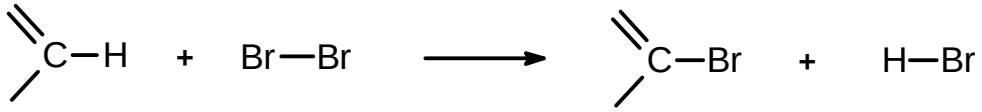


Figure 3-16. Bond rearrangement in the reaction of phenol with Br_2.

The following restrictions were imposed onto the carbon atom: it has to be part of an aromatic system, where an oxygen or nitrogen atom has to be conjugated to in a distance of two or four bonds. These constraints are valid for carbocyclic aromatic systems. To also use the reaction rule for heteroaromatic systems, the constraints would have to be adapted.

In order to make quantitative predictions, mechanisms for the estimation of the relative rates of bromination at the various positions of phenol have to be given. The following observations were used: Bromination of phenol gives about 80% p-bromophenol and 20% o-bromophenol,[26] allowing the conclusion that bromination in para-position is eight times faster than bromination in ortho-position (there are two ortho-positions !). As no absolute second order rate constants were available, bromination in ortho-position was set to 0.01 l/s•mole and in para-position to 0.08 l/s•mole. In order to account for the influence of a bromine substituent onto the rate of further bromination recourse was made to the following observation: The rate of nitration of bromobenzene is 3% of the rate of nitration of benzene.[27] It was therefore assumed that with each bromine substituent also the rate of bromination drops to 3% the rate without this additional bromine substituent.

With these rate constants the integration of the differential equations in the kinetic modeling was performed by the Gear algorithm.[21] Figure 3-17 shows the sequence of reaction products obtained in this reaction modeling. Figure 3-18 reproduces the results of the kinetic modeling of this system of reactions.

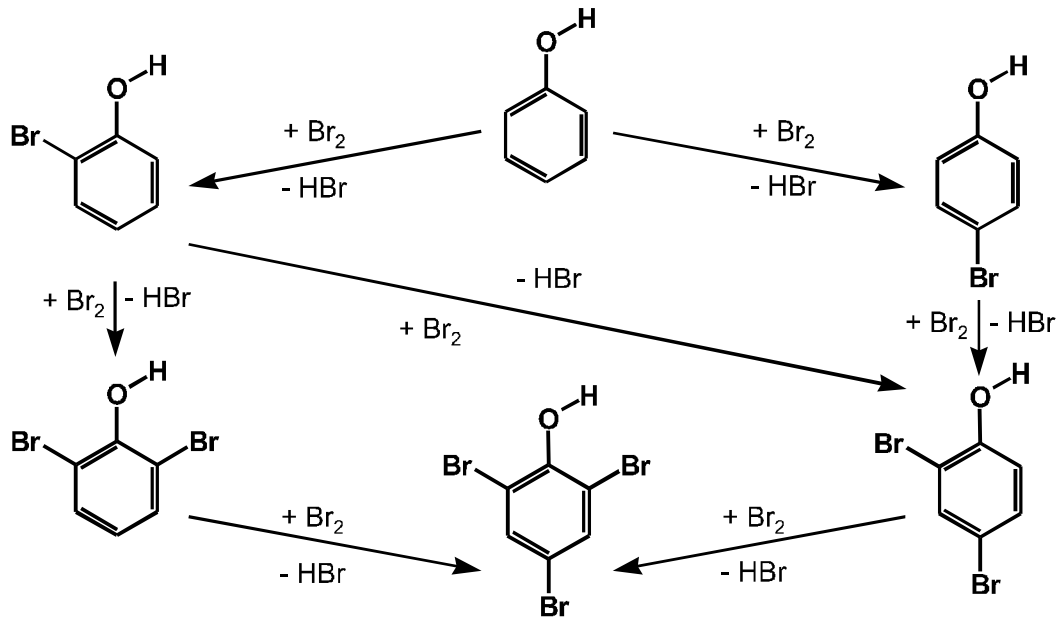


Figure 3-17. Reaction scheme of the bromination of phenol.

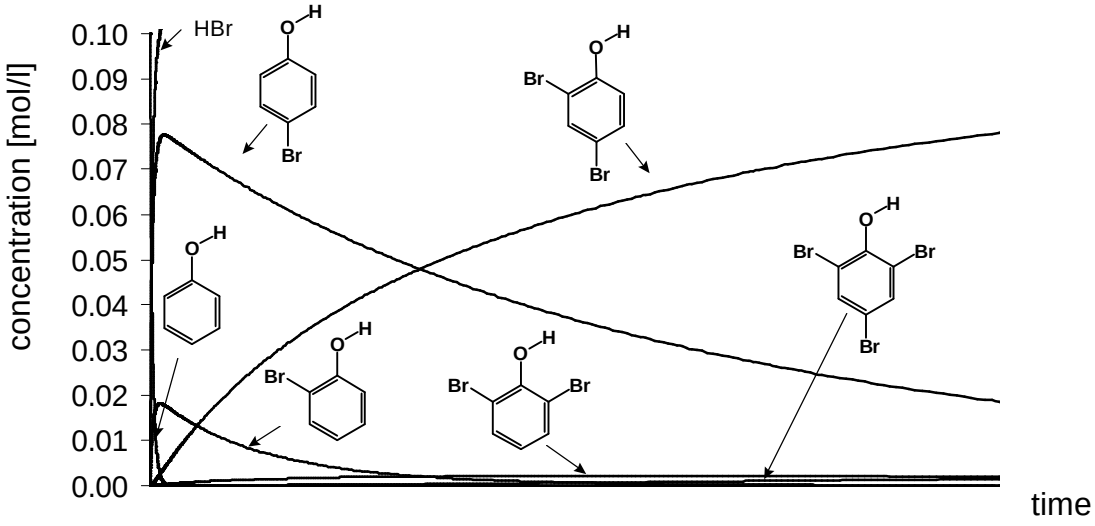


Figure 3-18. Concentration vs. time plot of the bromination of phenol.

Main features of the reaction rule file:

Rule header:

- number of reactors: 1
- number of phases: 1
- mode of phase 1: MIX
- kinetic of reactor 1: gear
  minimal concentration: 0.0
  maximal reaction time in sec: 1.e4
  conversion limit: 1.00
  (reaction prediction and integration stops if conversion limit is reached; here: if all reactions are generated (100%), or the integration exceeds the maximal reaction time, which will occur in our example)
- initial concentrations:

phenol: 0.1 mol/l
bromine: 0.4 mol/l

Reaction rules:

Reaction type 1:

Bromination (see Figure 3-16)
Deactivating substituents are not taken into account.

Reaction Generator:
Reaction levels: 3
Trace level: 0

Reaction rule file and input file:

The reaction rule and the CTX-input are contained as files (brphen.tcl and brphe.ctx) on the CD-ROM for the distribution of the EROS system.

3.5 Degradation of Chemicals in the Environment: Degradation of Atrazine and Prometon

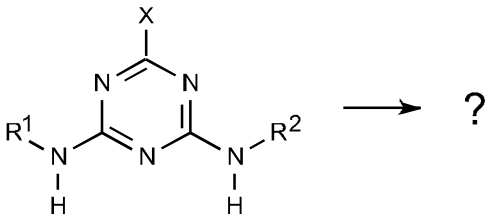


Figure 3-19. Degradation of s-triazines

Essentials to learn:

- how to derive rates for evaluating reaction rates from half-life times
- exhaustive exploration of degradation products
- how to deal with the multiple application of a chemical into the environment

Description:

The major degradation reactions of s-triazine herbicides such as prometon or atrazine under anaerobic conditions in soil are reductive dealkylation and hydrolysis. The concentration of the herbicides is generally at rather low concentration so that no reaction between the triazines will occur. Thus, no reactions of this chemical with other molecules of its kind have to be considered, but only those with chemicals having high concentration in the environment such as water or oxygen, or, in this case, bacteria that perform reductive dealkylation which is modeled by using hydrogen as a starting material. The general specifications for modeling these reactions were therefore: one reactor, one phase in the mode: monomolecular (which, in this case, corresponds to a pseudo-monomolecular process).

The two reaction types shown in Figure 3-20 were included in the knowledge base. Figure 3-21 shows the sequence of degradation reactions obtained for s-triazine herbicides with these two reaction types.

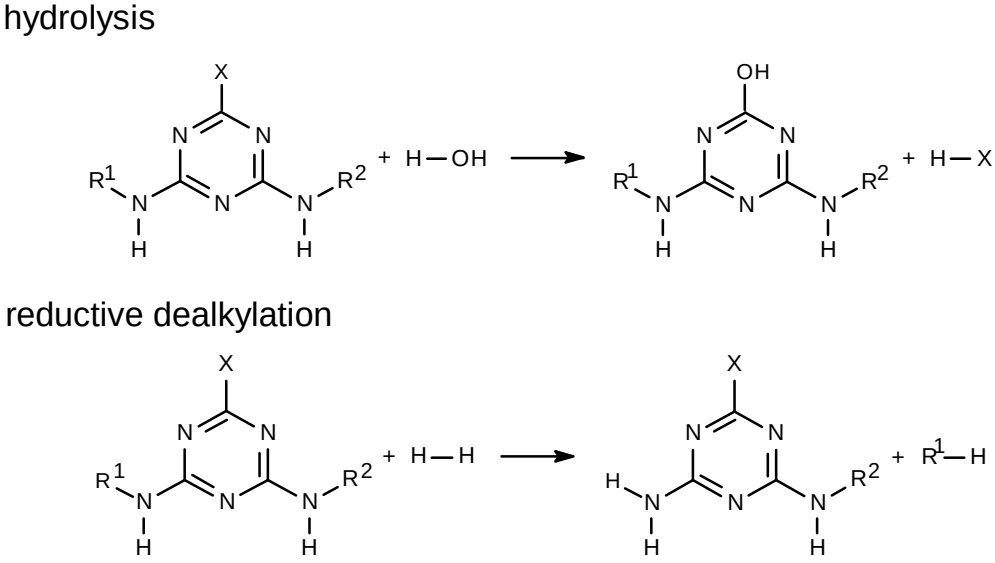


Figure 3-20. Degradation reactions of s-triazines in soil.

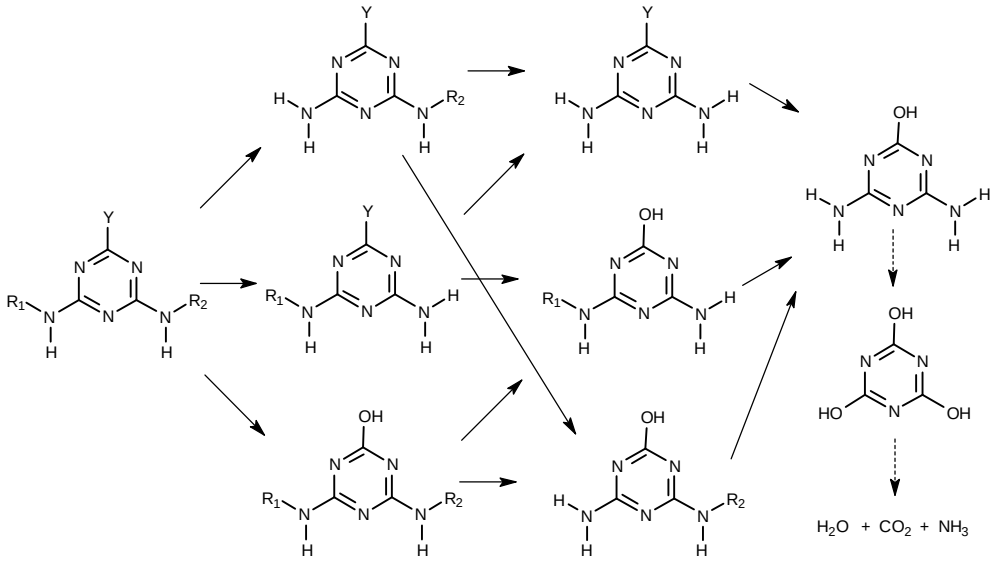


Figure 3-21. Reaction scheme of the degradation of s-triazines in soil.

The concentration dependence of the materials produced in this sequence of steps of pseudo first order rates is shown in Figure 3-22.

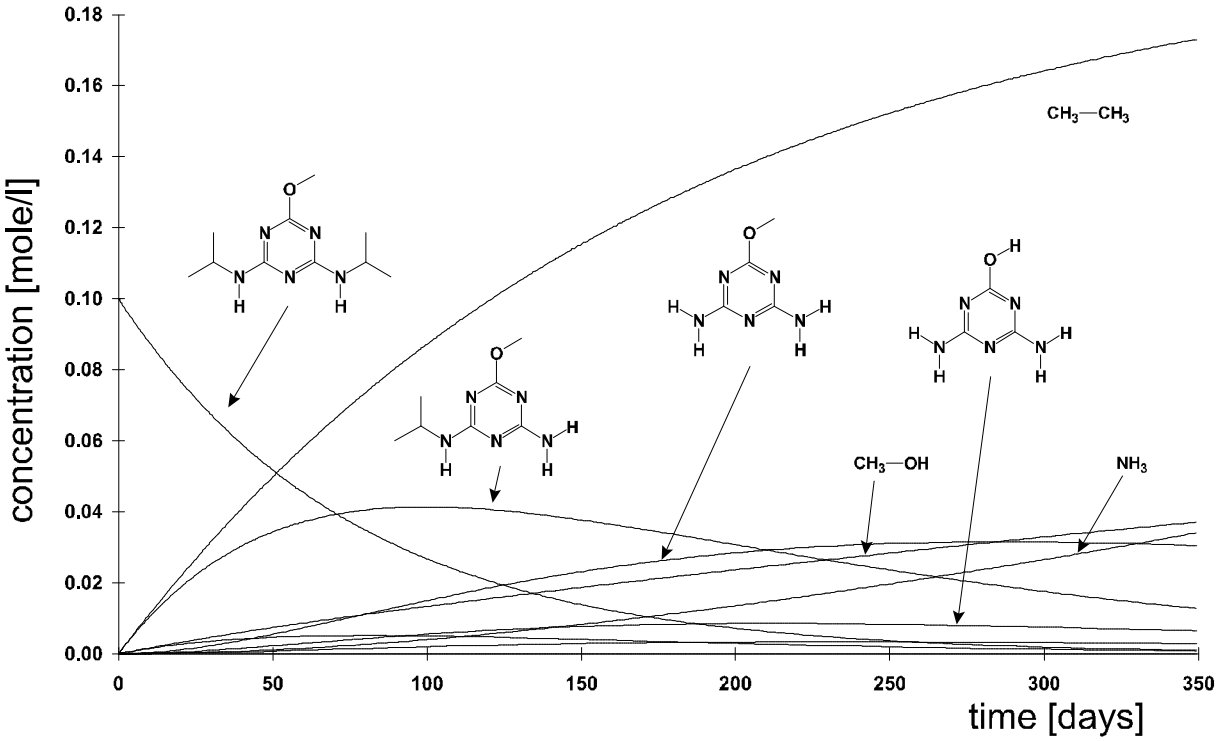


Figure 3-22. Concentration vs. time plot of the degradation of prometon.

a) Single application of a herbicide

Main features of the reaction rule file:

Rule header:

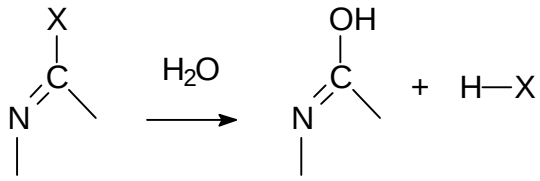
- number of reactors: 1
- number of phases: 1
- mode of phase 1: MONOMOLEC

- kinetic: gear
  minimal concentration: 0.0
  maximal reaction time in sec: 3.1536.e7 (one year)
  conversion limit of the starting materials: 0.99
  The integration is stopped when the maximal reaction time is reached or before, in the case of reaching the conversion limit of the starting materials
- initial concentrations:
  herbicide: 0.1 mol/l
  H_2O: 1.0 mol/l
  H_2 (as representation for the reductive equivalents): 1.0 mol/l

Reaction rules:

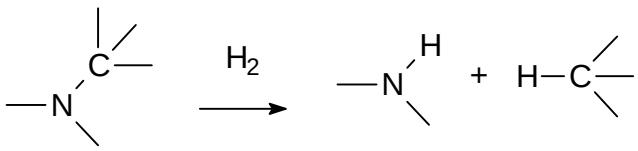
Reaction type 1 in RULE_1:

Hydrolysis



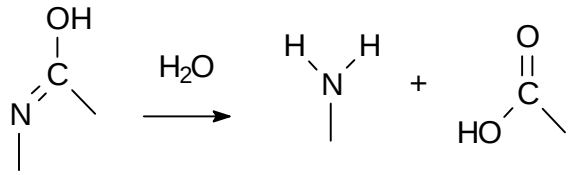
Reaction type 2 in RULE_2:

Reductive dealkylation



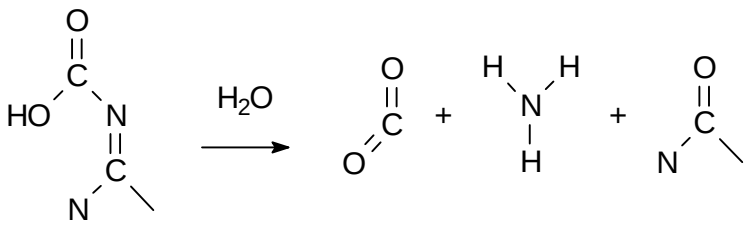
Reaction type 3 in RULE_3:

Hydrolysis of cyanuric acid



Reaction type 4 in RULE_4:

Decarboxylation



Reaction Generator:

Reaction levels: 9
Trace level: 0

Reaction rule file and input file:
The reaction rule and the CTX-input files are contained as files (triazine.tcl, atraz.ctx and promet.ctx) on the CD-ROM for the distribution of the EROS system.

Note: The displayed reaction time in the concentration-time plot may be less than the wished reaction time (see chapter 5.3).

b) Multiple application

This example also shows how to deal with reactions of chemicals that are regularly applied such as drugs or, as with this example, plant protection agents. It is assumed that prometon is regularly applied once a year. The development of the degradation products over time is shown in Figure 3-23.

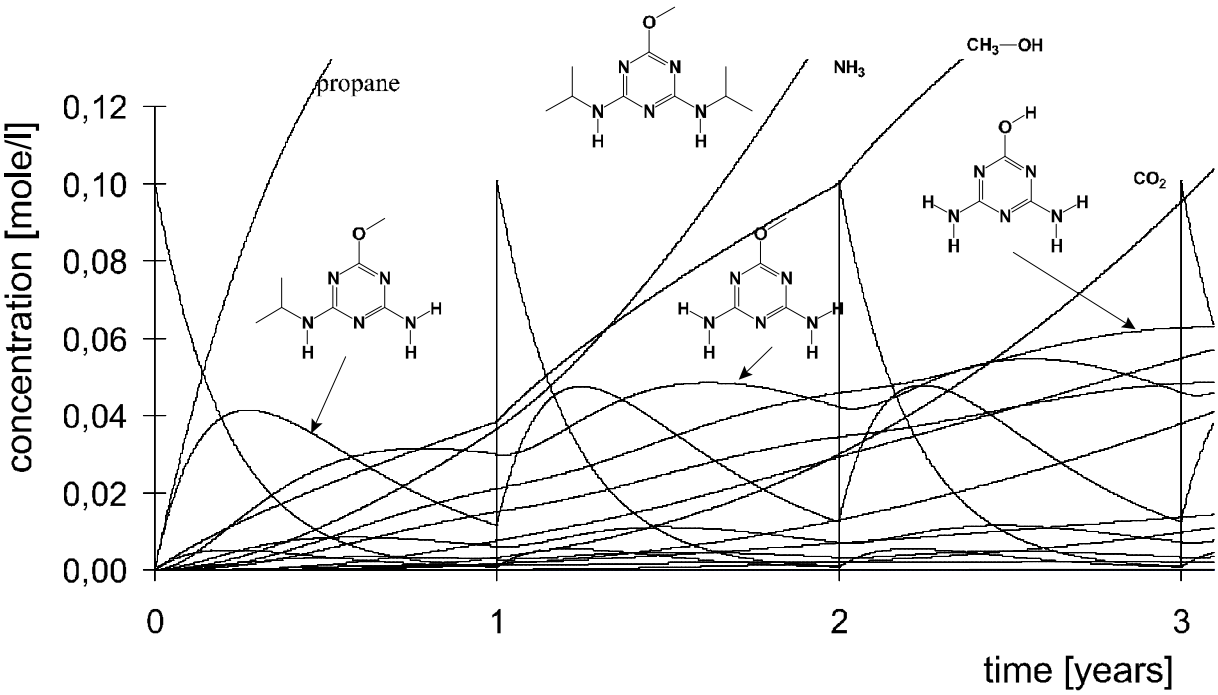


Figure 3-23. Concentration vs. time plot of the degradation of prometon applied every year.

Main features of the reaction rule file:

If you transfer a variable named multi to the rules (e.g.: multi#1, the value will not be recognized) the application of the triazine is enabled every year. Together with this the reaction time is set to 1.10376e8 seconds (3.5 years) and the conversion limit is set to 1.01, so that it cannot be reached at all.
In both cases, single application (without multi) and multiple application (with multi#1), you can specify the reaction time and the initial concentration of the triazine, which is added every year in the case of multiple application.
Therefore set the concentration e.g. with conc:0.001 with the advanced options. Due to a slight inconsistency in the Tcl interface (plus signs are not accepted in a number; see chapter

5.1.4) and the fact that Tcl does not have different data types, specify huge numbers as character strings, which are given to the rules. This is true for the reaction time (rxtime \$2.1e8). The backslash in front of the $ sign is necessary because of the shell wrapper which is called from the GUI. Additionally you can set the conversion limit in the case of a single application with e.g. conv_lim:0.5.

3.6 Simulation of Mass Spectra

The mass spectrum of N-methyl-N-propyl-butylamine

Essentials to learn:

- How to simulate mass spectra
- The reaction rules for mass spectra simulation
- The evaluation of a reaction network based on the probability of reaction steps

Description:

The mass spectra that can be simulated with EROS7 refer to 70eV EI mass spectra. This is so because the evaluation mechanisms in EROS7 have been derived from a database of 70eV EI mass spectra. At present, the quality of the simulated mass spectra is, in most cases, still quite unsatisfactory. This is mainly due to the limited set of reaction types included in the rule file. In fact, apart from ionization, only four fragmentation types (Figure 3-24) are included in the rule file.

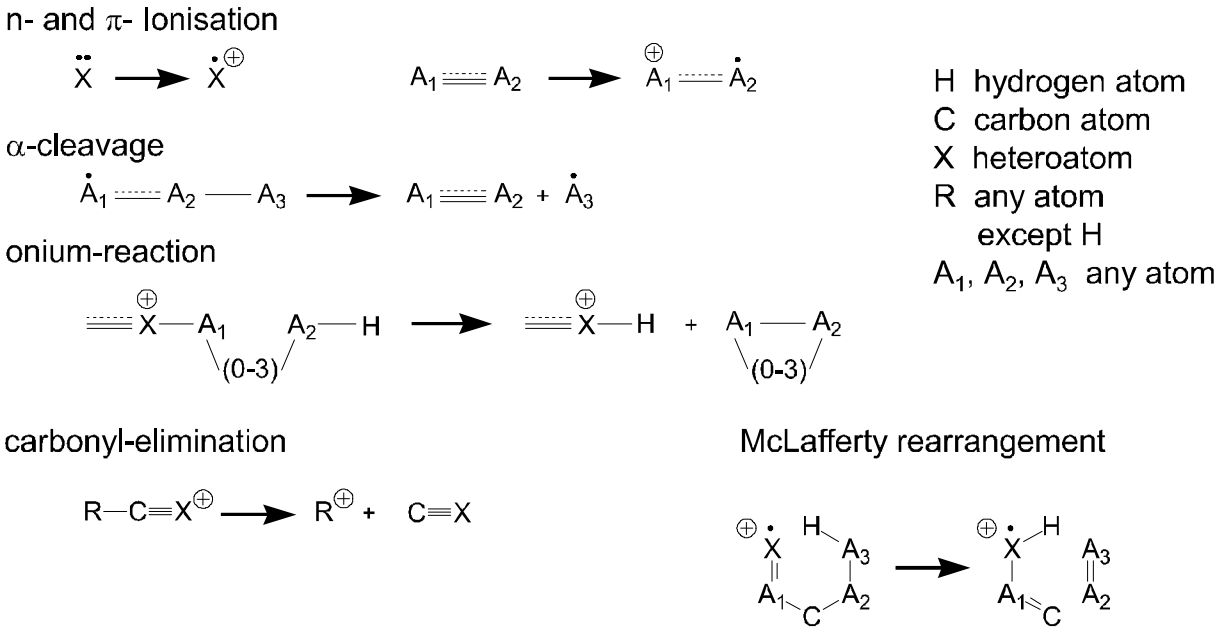


Figure 3-24. Fragmentation types included in the rule file.

These fragmentation reactions are mainly valid for aliphatic compounds; specific fragmentation reactions of aromatic systems are conspicuously absent. Even for aliphatic systems only the major fragmentation reactions are considered; some important reactions such as hydrogen rearrangements or inductive cleavage are not yet accounted for. The reason is that good evaluation mechanisms for hydrogen rearrangements have not yet been developed. Even for the fragmentation types shown in Figure 3-24, the evaluation is based on work that has been performed five years ago.[5]

Having said this, it is nevertheless true that for many monofunctional aliphatic compounds the major peaks in the mass spectrum can be reproduced. Even the peak intensities often are quite in agreement with experiment giving support to the mathematical basis for calculating peak intensities on the basis of estimations of probabilities for the individual fragmentation steps.

The example discussed here deal with the simulation of the mass spectrum of a simple monofunctional aliphatic compound, the trialkylamine N-methyl-N-propyl-butylamine.

The resulting simulated mass spectrum is shown in Figure 3-25.

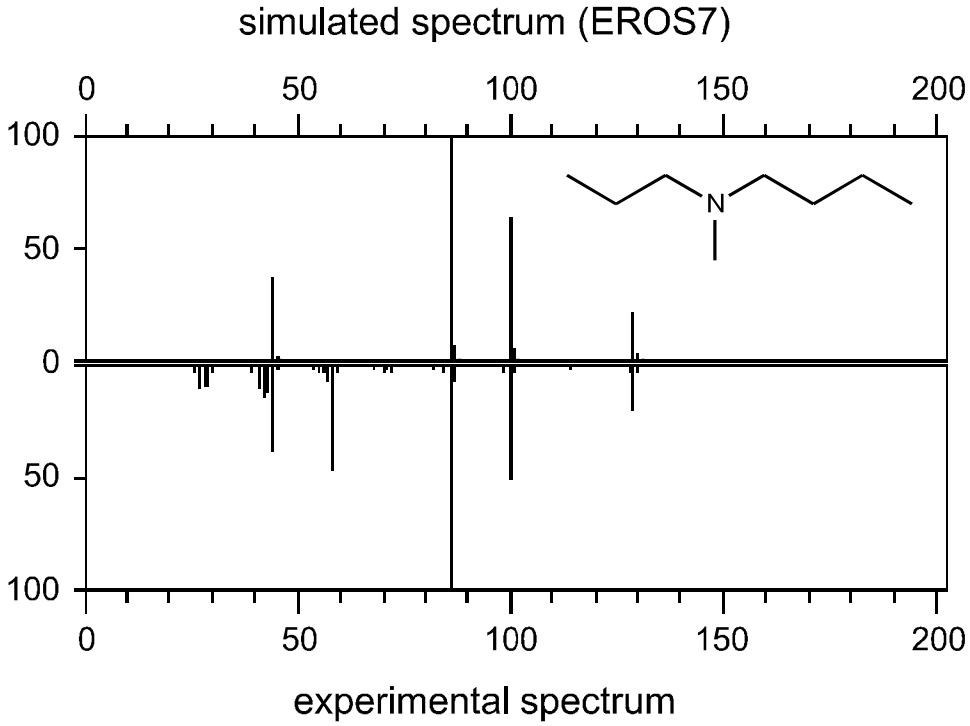


Figure 3-25. Simulated and experimental mass spectrum of N-methyl-N-propyl-butylamine.

If you switched on writing the file with the mass spectra, an additional button will appear to view the simulated mass spectra in the result section of the EROS GUI.

Main features of the reaction rule file:

Rule header:

- number of reactors: 1
- number of phases: 1
- mode of phase 1: MONOMOLEC
- kinetic: prob_kin
  minimal concentration: 1.e-5

Reaction rules:

Reaction type 1 in RULE_1:
Ionization (see Figure 3-24)

Reaction type 2 in RULE_2:
Alpha cleavage (see Figure 3-24)

Reaction type 3 in RULE_3:
Onium reaction (see Figure 3-24)

Reaction type 4 in RULE_4:
Carbonyl elimination (see Figure 3-24)

Reaction type 5 in RULE_5:
McLafferty reaction (see Figure 3-24)

Reaction Generator:
Reaction levels: more than 3
Trace level: 0

Advanced Options:
In this example it is possible to use an additional reaction type by switching on hydrogen rearrangements with a row reactivity specification. It does not take the differences due to the size of the molecule in account. Click on the Advanced Options Button in the Data Input area of EROS7 - User Interface (EROS - GUI) and pass a variable called huml with the value 1 (type 'huml#1') to the rule file. (Reaction levels: more than 9).
Attention: If you switch on the hydrogen rearrangement many more reactions are simulated ans so the EROS run take quite longer!

Data Output:
Mass spectra: Name of the file that should contain the mass spectra (only produced, if the MS file is switched on).

Reaction rule file and input file:
The reaction rule and the CTX-input are contained as files (ms.tcl and ms1.ctx) on the CD-ROM for the distribution of the EROS system.

3.7 Working with Two Reactors: Acetal Formation from Propanal and Ethanol and Subsequent Simulation of the Mass Spectra

Essentials to learn:

- Working with two reactors
- Combining a laboratory reaction with the simulation of mass spectra

Description:

This example serves to illustrate how a combination of two reactors can be used. The first reactor performs a laboratory reaction, the second models the events in a mass spectrometer. The reaction chosen is the formation of an acetal from propanal and ethanol (Figure 3-26).

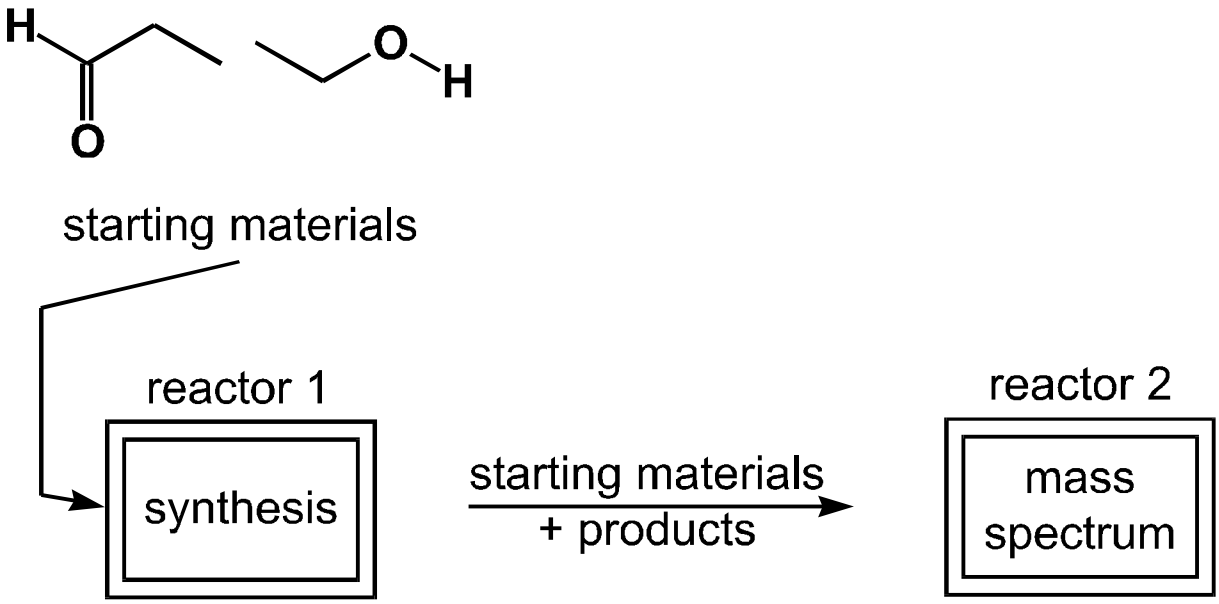


Figure 3-26. Formation of an acetal and simulation of a mass spectrum.

The following reaction types have been coded for this synthesis.

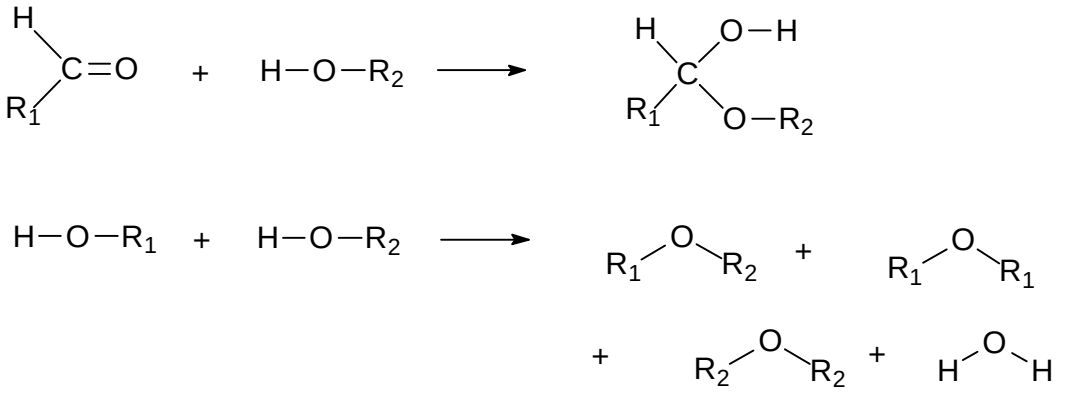


Figure 3-27. Reaction types in the rule file for the first reactor.

Each reaction product (and each starting material) is individually handed over to the second reactor in order to have its mass spectrum simulated (see Figure 3-26).

The reaction types used for the simulation of mass spectra are the same as in the previous example (see Figure 3-24).

The sequence of reaction steps generated in the first reactor is shown in Figure 3-28.

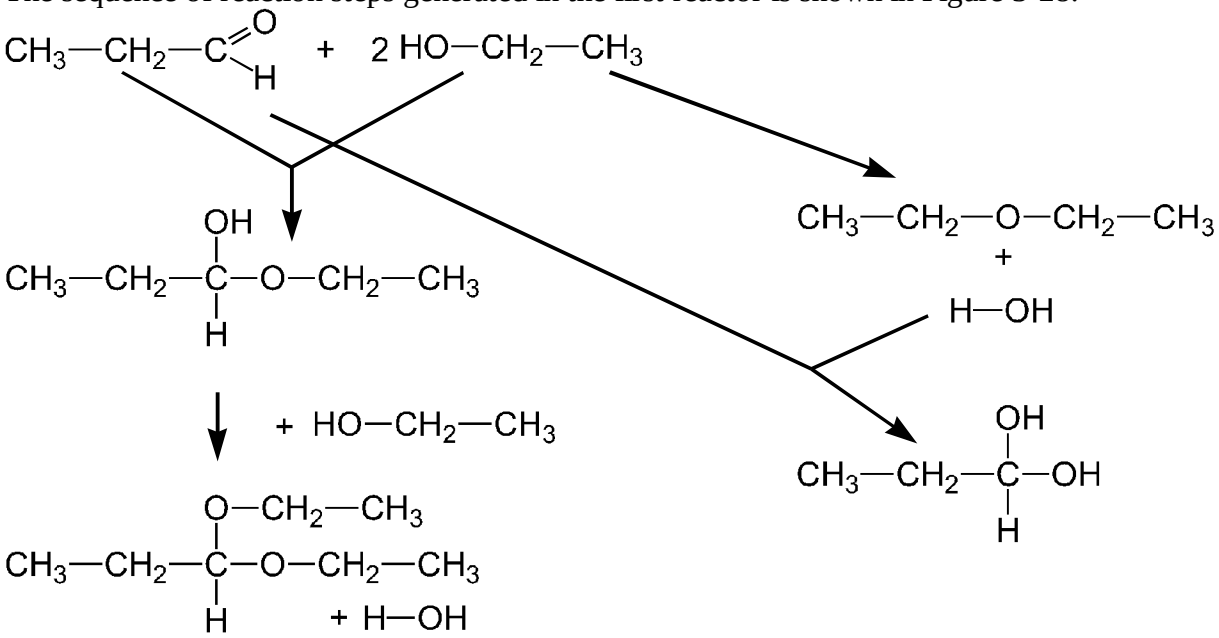


Figure 3-28. Sequence of reaction steps in the acetal formation.

The mass spectra simulated for the two starting materials are shown in Figure 3-29 and compared with their experimental mass spectra as taken from the MassLib spectra database.[28]

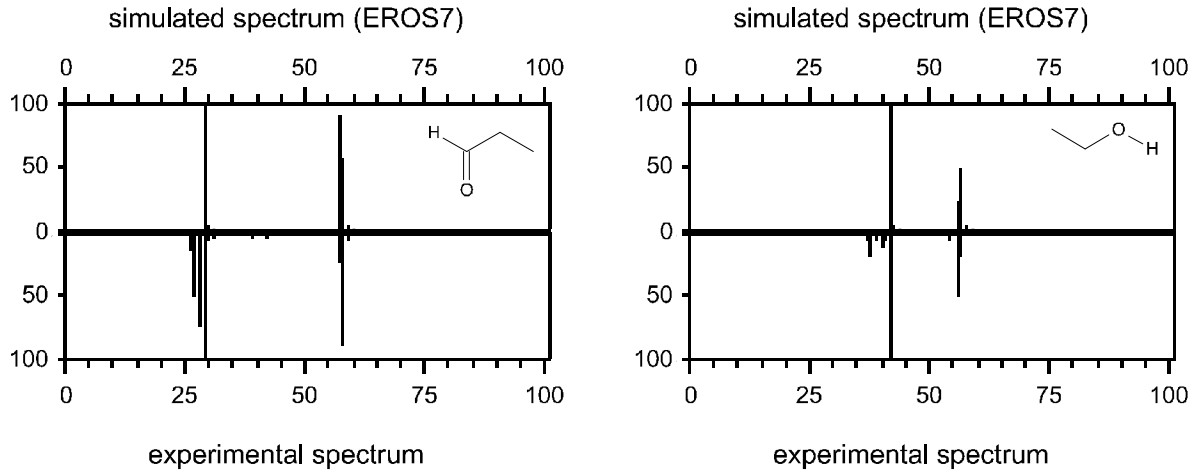


Figure 3-29. Mass spectra of starting materials.

The comparison of the simulated with the experimental [28] mass spectra of the products in this reaction sequence is made in Figure 3-30.

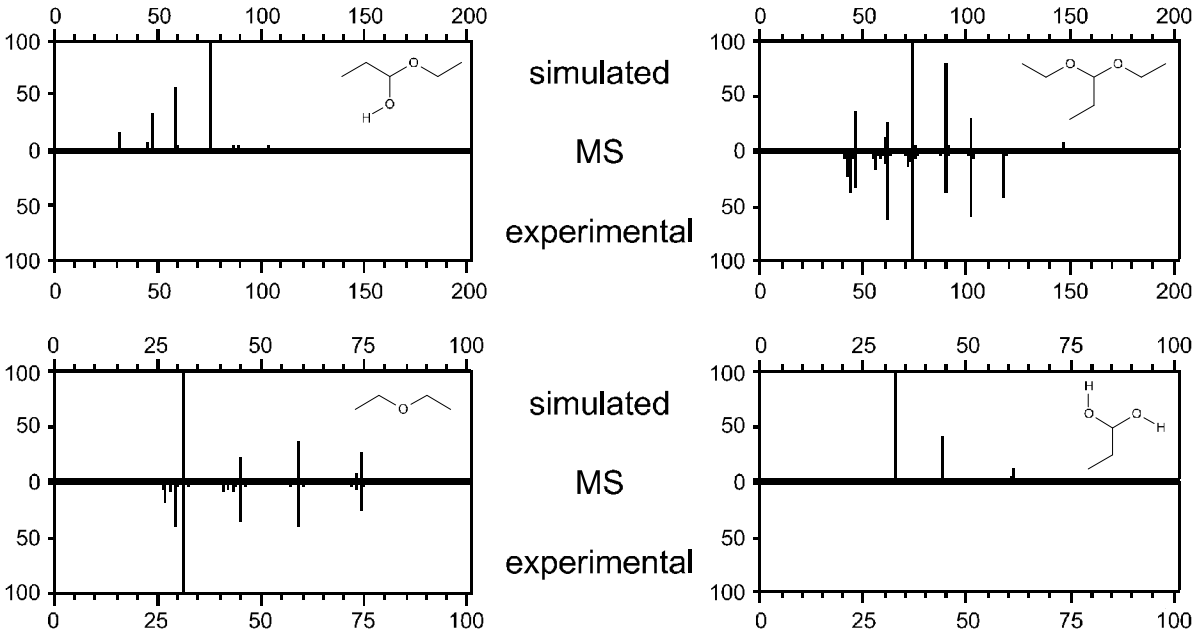


Figure 3-30. Mass spectra of products.

Note, that the experimental mass spectra are not included on the CD.

Main features of the reaction rule file:

Rule header:

- number of reactors: 2
- number of phases in the first reactor: 1
- number of phases in the second reactor: 1
- mode of phase 1 in reactor 1: MIX
- mode of phase 2 in reactor 2: MONOMOLEC
- kinetic in reactor 1: gear
  minimal concentration: 1.0e-5
  reaction time: 3600.
  conversion limit: 0.95
- kinetic in reactor 2: prob_kin

Reaction rules:

Reaction type 1 in RULE_1:
Ionization (see Figure 3-24)

Reaction type 2 in RULE_2:
Alpha cleavage (see Figure 3-24)

Reaction type 3 in RULE_3:
Onium reaction (see Figure 3-24)

Reaction type 4 in RULE_4:
Carbonyl elimination (see Figure 3-24)

Reaction type 5 in RULE_5:
McLafferty reaction (see Figure 3-24)

Reaction type 6 in RULE_7:
Formation of the acetal (see Figure 3-27)

Reaction Generator:
Reaction levels: 5
Trace level: 0

Advanced Options:
In this example it is also possible to use an additional reaction type by switching on hydrogen rearrangement. Click on the Advanced Options Button in the Data Input area of EROS7 - User Interface (EROS - GUI) and pass a variable called huml with the value 1 (type ‘huml#1’) to the rule file. (Reaction levels: more than 10).

Data Output:
Mass spectra: Name of the file that should contain the mass spectra

Reaction rule file and input file:
The reaction rule and the CTX-input are contained as files (mixms.tcl and mixms.ctx) on the CD-ROM for the distribution of the EROS system.

4 Writing Your Own Reaction Rules

This section of the manual is still in preparation.
A manual in German language for writing reaction rules in *Tcl* (or *C++*) is contained as Appendix in the PhD thesis of Dr. Robert Höllering, University Erlangen - Nürnberg. This thesis is available over the internet at http://vermeer.organik.uni-erlangen.de/dissertationen/data/dissertation/Robert_Hoellering/html/ .

5 Known Limitations and Further Development

It should be realized that EROS is still in active development, particularly as concerns the knowledge base. Thus, there are a number of problems and limitations that we know of and that we are trying to resolve. It is important for the users and particularly the system managers of EROS to seek contact with us in order to be provided with the most recent developments.

5.1 Reaction Rules

5.1.1 Number of Reaction Rules

The singly most disturbing limitation is that the number and the scope of the reaction types available as coded reaction rules is still rather limited. The ones that are presently available have mainly been coded to show by spotlights the range of applications open to EROS.
Thus, the coverage of organic reactions presently is rather fragmentary.
In this situation, the user may code her/his own reaction rules. The present set-up requires a knowledge of the scripting language *Tcl* which can only be expected from experienced users or system managers.

5.1.2 Automatic Rule Generation

Since many years we have been working on automatic reaction rule generation. The previous version, EROS 6, had such a facility. Since then we have applied machine learning techniques such as conceptual clustering methods to perceive reaction types and thus extract knowledge from reaction databases.[29] The system HORACE was based on topological and physicochemical criteria to perform this task.
Our present answer to this problem are self-organizing neural networks such as the one proposed by Kohonen. Methods have been developed for the clustering of reactions extracted from databases into landscapes of reactions that allow the perception of reaction types.[30][31][32]
This has been further developed to explore the feasibility and scope of a reaction type. Mechanisms have already been provided for accessing a Kohonen network from a reaction rule. Thus, we are on the way to automatically extract knowledge on chemical reactions from reaction databases and make such knowledge available to EROS.

5.1.3 Number of Neural Networks

The evaluation of reaction types can be supported by neural networks such as backpropagation (BPG) networks, Kohonen networks, or counterpropagation (CPG) networks. For each rule written in *Tcl* only one BPG network, and one Kohonen/CPG network can be provided. The initialization of the network has to be made in the RULE_INFO part of the reaction rule that uses such a network.

5.1.4 Floating Point Values and Integer Number

Floating point values and integer numbers which are transferred to the core system are not allowed to contain +-signes. For one problem and how to avoid it see chapter 3.5.

5.2 Number of Intermediates and Products

The current version of EROS keeps all reaction intermediates and products in memory and does not save the molecules on disk when the required storage space exceeds memory size. Storage requirements may become particularly critical in the simulation of experiments of combinatorial chemistry when a large number of products is generated. The memory necessary for an EROS run depends on the number and size of different structures produced and on the contents of the reaction rules such as number of rules and usage of physicochemical variables. The EROS 7 system itself needs slightly more than 20 MB. Dependent on the size of the molecules and the physicochemical variables used about 5 - 250 kB are needed per molecule.
Presently, care has to be taken that enough memory is available. Otherwise, EROS 7 may crash uncontrollably. When the development of EROS was initiated, not all *C++* compilers had the exception handling implemented and thus a check for insufficient memory at important steps in the program could not be made.
The next version of EROS will provide for an automatic save onto disk if the number of molecules exceeds a given threshold. Thus, this limitation will have been overcome.

5.3 Integration of Kinetics

It is recommended to use the GEAR algorithm for the integration of the differential equations. The other integration methods (Runge-Kutta and Runge-Kutta-Merson) may fail in some cases. The multi-time dosage of starting materials works only with the GEAR algorithm, because the other integration methods are less stable.

The file with the data for the concentration-time curves is presently only written with the GEAR algorithm. The end concentrations are calculated by all three integration methods. If minimal_concentration is set to 0.0, the concentration values are written into the file PS1.prd. If minimal_concentration is set to a value higher than 0.0, a file PS#.prd is written for each reaction level, where # is the number of the level. The concentration values for the entire reaction network are in the file PS#.prd, where # is the highest number of reaction levels. All these files have the same file format. The name of the files currently cannot be changed.

The files PS#.prd are column oriented. The first column gives the time in seconds followed by columns for molecule 0, molecule 1, and so on until the last molecule. The molecule numbers are the same as in the structure file. Because all molecules are stored with numbers starting with zero and are copied to new molecule numbers before they take part in a reaction, the concentration values for molecule 0 and, if you start your simulation with more than one molecule, the next ones are all 0.0, too.

The total reaction time in these files may be less than the chosen reaction time because these files are limited to 4999 time intervals of the integration. The size of the intervals is chosen by an internal algorithm, which may exceed 4999. In this case the concentration values end for all compounds at that time, where the maximum of 4999 intervals is reached.

5.4 Physicochemical Variables

Most of the physicochemical variables are still calculated from a connection table representation of molecules. The structures in the MOSES format are automatically converted to one reasonable connection table. The calculation of physicochemical descriptors is initiated with this structure and does therefore still suffer from the limitations of a connection table. In particular, it cannot be controlled which mesomeric structure will be generated in the conversion of a MOSES representation.

5.5 Manual

The manual still needs extensions in various chapters. In particular, the chapter 4, Writing Your Own Reaction Rules, still has to be written in English. A German version is contained in the dissertation by Dr. Robert Höllering, University Erlangen-Nürnberg, 1998 which can be accessed over the internet at http://vermeer.organik.uni-erlangen.de/dissertationen/data/dissertation/Robert_Hoellering/html/ (A few functionality described here is not yet working: combine_elsys, some group handling functionality, and the handling of the internal error flag).
This dissertation is a rich source of additional information on the EROS system. However, all this information is in German.
Detailed information on the MOSES data structure and its implementation in *C++* can be obtained from the dissertation of Dr. Susanne Bauerschmidt, which can also be accessed online at http://vermeer.organik.uni-erlangen.de/dissertationen/data/dissertation/Susanne_Bauerschmidt/html . Again, however, all this information is in German.

6 References

[1] EROS - A Computer Program for Generating Sequences of Reactions
J. Gasteiger, C. Jochum
Topics Curr. Chem. *74*, 93-126 (1978)

[2] A New Treatment of Chemical Reactivity: Development of EROS, an Expert System for Reaction Prediction and Synthesis Design
J. Gasteiger, M. G. Hutchings, B. Christoph, L. Gann, C. Hiller, P. Löw, M. Marsili, H. Saller, K. Yuki,
Topics Curr. Chem., *137*, 19-73 (1987)

[3] Computergestützte Planung organisch-chemischer Synthesen: Die zweite Programmgeneration
W.-D. Ihlenfeldt, J. Gasteiger
Angew. Chem., *107*, 2807-2829 (1995)

Computer-Assisted Planning of Organic Syntheses: The Second Generation of Programs
W. D. Ihlenfeldt, J. Gasteiger
Angew. Chem. Int. Ed. Engl., *34*, 2613-2633 (1995)

[4] Automated Derivation of Reaction Rules for the EROS 6.0 System for Reaction Prediction
P. Röse, J. Gasteiger
Anal. Chim. Acta, *235*, 163-168 (1990)

[5] Prediction of Mass Spectra from Structural Information
J. Gasteiger, W. Hanebeck, K.-P. Schulz
J. Chem. Inf. Comput. Sci., *32*, 264-271 (1992)

[6] The Simulation of Organic Reactions: From the Degradation of Chemicals to Combinatorial Synthesis
R. Höllering, T. Kostka, J. Gasteiger, L. Steinhauer, K.-P. Schulz, in preparation

[7] Iterative Partial Equalization of Orbital Electronegativity - A Rapid Access to Atomic Charges
J. Gasteiger, M. Marsili
Tetrahedron *36*, 3219-3228 (1980)

[8] Berechnung der Ladungsverteilung in konjugierten Systemen durch eine Quantifizierung des Mesomeriekonzeptes
J. Gasteiger, H. Saller
Angew. Chem. *97*, 699-701 (1985)

Calculation of the Charge Distribution in Conjugated Systems by a Quantification of the Resonance Concept
J. Gasteiger, H. Saller
Angew. Chem. Intern. Ed. Engl. *24*, 687-689 (1985)

[9] Residual Electronegativity - An Empirical Quantification of Polar Influences and its Application to the Proton Affinity of Amines
M. G. Hutchings, J. Gasteiger
Tetrahedron Lett. *24*, 2541-2544 (1983)

[10] A. Fröhlich, TU München, 1993, unpublished results

[11] Quantification of Effective Polarisability. Applications to Studies of X-Ray Photoelectron Spectroscopy and Alkylamine Protonation
J. Gasteiger, M. G. Hutchings
J. Chem. Soc. Perkin *2*, 1984, 559-564

[12] Dipole Moments Obtained by Iterative Partial Equilisation of Orbital Electronegativity

J. Gasteiger, M. D. Guillen
J. Chem. Research (S) 1983, 304-305; (M) 1983, 2611-2624

[13] Quantitative Models of Gas-Phase Proton Transfer Reactions Involving Alcohols, Ethers, and their Thio Analogs. Correlation Analyses Based on Residual Electrone-gativity and Effective Polarizability
J. Gasteiger, M. G. Hutchings
J. Amer. Chem. Soc. *106*, 6489-6495 (1984)

[14] A Quantitative Empirical Treatment of C-13 NMR Chemical Shift Variations on Successive Substitution of Methane by Halogen Atoms
J. Gasteiger, I. Suryanarayana
Magn. Reson. Chem. *23*, 156-157 (1985)

[15] A Quantitative Description of Fundamental Polar Reaction Types. Proton and Hydride Transfer Reactions Connecting Alcohols and Carbonyl Compounds in the Gas Phase
M. G. Hutchings, J .Gasteiger
J. Chem. Soc. Perkin *2*, 1986, 447-454

[16] Correlation Analyses of the Aqueous Phase Acidities of Alcohols and Gem-Diols, and of Carbonyl Hydration Equlibria, using Electronic and Structural Parameters
M. G. Hutchings, J. Gasteiger
J. Chem. Soc. Perkin *2*, 1986, 455-462

[17] Elucidating Chemical Reactivity by Pattern Recognition Methods
J. Gasteiger, H. Saller, P. Löw
Anal. Chim. Acta, *191*, 111-123 (1986)

[18] Empirical Methods for the Calculation of Physicochemical Data of Organic Compounds
J. Gasteiger
in: "Physical Property Prediction in Organic Chemistry"
Editor: C. Jochum, M. G. Hicks, J. Sunkel
Springer-Verlag, Heidelberg, 1988, S. 119-138

[19] Tcl and the Tk Toolkit
J. K. Ousterhout
Adison-Wesley, 1984

[20] Computer Assisted Prediction of the Degradation of Chemicals: Hydrolysis of Amides and Benzoylphenylureas
J. Gasteiger, U. Hondelmann, P. Röse, W. Witzenbichler
J. Chem. Soc. Perkin Trans. *2*, 1995, 193-204

[21] Numerical Initial Value Problems in Ordinary Differential Equations
C. W. Gear
Prentice Hall, Englewood Cliffs, N.J. 1971

[22] Numerical Recipies, The Art of Scientific Computing
W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vettering
Cambrige University Press, 1989

[23] Overcoming the Limitations of a Connection Table Description: A Universal Representation of Chemical Species
S. Bauerschmidt, J. Gasteiger
J. Chem. Inf. Comput. Sci., *37*, 705-714 (1997)

[24] Synthesis and Evaluation of Three 1,4-Benzodiazepine Libraries
B. A. Bunin, M. J. Plunkett, J. A. Ellman
in: Combinatorial Peptide and Nonpeptide Libraries, A Handbook, G. Jung (Ed.)
VCH Verlagsgesellschaft, Weinheim, 1996

[25] Hash Codes for the Identification and Classification of Molecular Structure Elements

W.-D. Ihlenfeldt, J. Gasteiger
J. Comput. Chem., *15*, 793-813 (1994)

[26] Introduction to Organic Chemistry
A. Streitwieser Jr., C. H. Heathcock
Macmillan Publishing Company, New York, 3. Edition, 1985

[27] Grundlagen der Organischen Chemie
H. R. Christen
Verlag Diesterweg-Salle, Frankfurt a. M., 2. Auflage, 1972

[28] MassLib vom MPI für Kohleforschung, Mülheim a. d. Ruhr

[29] Extraction of Chemical Knowledge from Organic Reaction Data by Automatic Hierarchical Classification and Generalization
L. Chen, J. Gasteiger, J. R. Rose
in: Software-Entwicklung in der Chemie *9*, R. Moll (Ed.)
GDCh, Frankfurt a. M., 1995

[30] Knowledge Discovery in Reaction Databases: Landscaping Organic Reactions by a Self Organizing Neural Network
Lingran Chen, J. Gasteiger
J. Am. Chem. Soc., *119*, 4033-4042, (1997)

[31] Organische Reaktionen mit Hilfe neuronaler Netze klassifiziert: Michael-Additionen, Friedel-Crafts-Alkylierungen durch Alkene und verwandte Reaktionen
L. Chen, J. Gasteiger
Angew. Chem., *7*, 844-846, (1996)

Organic Reactions Classified by Neural Networks: Michael Additions, Friedel-Crafts Alkylations by Alkenes, and Related Reactions
L. Chen, J. Gasteiger
Angew. Chem. Int. Ed. Engl., *7*, 763-765, 1996

[32] Classification of Organic Reactions: Similarity of Reactions Based on Changes in the Electronic Features of Oxygen Atoms at the Reaction Sites
H. Satoh, O. Sacher, T. Nakata, L. Chen, J. Gasteiger, K. Funatsu
J. Chem. Inf. Comput. Sci., *38*, 210-219, 1998

7 Appendices

7.1 Manual for the CACTVS Editor csed

The manual for the CACTVS Editor csed is contained as html - file on the CD-ROM for the distribution of the EROS system.

7.2 Manual for the CACTVS Tree Tool cstr

The manual for the CACTVS Tree Tool cstr is contained as html - file on the CD-ROM for the distribution of the EROS system.